An empirical study on the Rasch Model

by

Mak Chiu

A thesis submitted to the
School of Education of the
Chinese University of Hong Kong
in partial fulfilment of
the requirements for the
Degree of Master of Arts in Education

July, 1975

This thesis, written by

MAK CHIU

under the supervision of the Thesis Committee, and approved by all its members, has been accepted by the Graduate School of the Chinese University of Hong Kong, in partial fulfilment of the requirements for the degree of

MASTER OF ARTS (EDUCATION)

Date 30th August, 1975

THESIS COMMITTEE

K. Cheng (Chairman)

Tam Tim Kin

SHIU L. KONG

C. K. Leung

ALBERT H. YEE

TABLE OF CONTENT

LIST OF TABLES

Table page

LIST OF ILLUSTRATIONS

Figure page

CHAPTER ONE
THE PROBLEM

Problem Statement

There are many models for the measurement and description of latent traits. They are all different in the number of parameters involved and the assumptions of the probability P_{ij} for a subject i giving a right response to an item j .

The abilities or the latent traits that a test item is intended to measure , may be of n dimensions.

$$\theta = (\theta_1 , \theta_2 , \theta_3 , \quad , \theta_n)$$

For all models mentioned here, only one dimensional ability is assumed (Lord,1967,p.366).

(1) Normal Ogive Model

The probability of right response, from a subject of ability θ , to an item g , of discrimination a_g and difficulty b_g is given by the following formula,

$$\begin{aligned} P_g(\theta) &= P_g(\theta,a_g,b_g) \\ &= \int_{-\infty}^{a_g(\theta-b_g)} \varphi(t)\,dt \\ &= \Phi\left[a_g(\theta-b_g)\right] \end{aligned}$$

where $\varphi(t)$ is the normal distribution function (Lord,1967,pp. 361-362).

(2) Modified Normal Ogive Model

In this model, guessing factor is also considered. The probability is given by

$$P_g(\theta) = c_g + (1-c_g)\,\Phi\, a_g(\theta - b_g)$$

where is the normal ogive function.
The probability function lies in the range from c_g to one. Usually for the multiple choice type test item of five alternatives, c_g is equal to 0.2 (Lord,1967,p.404).

(3) Lazarsfeld's Linear Model

This model has a very simple form of the probability function.

$$P_g(\theta) = a_g(\theta - b_g)$$

This model is designed for attitude investigation. The θ is restricted to an interval, in which all the values of $P_g(\theta)$ lie between zero and one(Lord,1967,p.404).

(4) Logistic Test Model

This model nearly coincides with the Normal Ogive Model. The probability function has the form of a Logistic Cumulative Distribution function.

$$P_g(\theta) = \frac{\exp(D\, a_g(\theta - b_g))}{1 + \exp(D\, a_g(\theta - b_g))}$$

or takes the following form

$$P_g(\theta) = \frac{1}{1 + \exp(-D\, a_g(\theta - b_g))}$$

where D=1.7 denotes a number that serves a unit scaling factor to maximize agreement between qualitative details in the Normal and Logistic Models(Lord,1967,p.404).

(4) Rasch Model

This is a restricted Logistic Model. Uniform discrimination is assumed.

Set a_g(discrimination) = 1

and D = 1

in the Logistic Model, then

$$P_g(\theta) = \frac{1}{1 + \exp(\,-(\theta - b_g))}$$

$$= \frac{1}{1 + \exp(\,b_g - \theta\,)}$$

Let $b_g^* = \exp(b_g)$

$\theta^* = \exp(\theta)$

$$P_g(\theta) = \frac{1}{1 + b_g^*/\theta^*}$$

$$= \frac{\theta^*/b_g^*}{1 + \theta^*/b_g^*}$$

(Lord,1967,p.402). According to Wright's notation(1969), the difficulty b_g^* is replaced by the item easiness of item j , E_j,

$$E_j = 1/b_g^*$$

and Θ^* is replaced by the person ability of the person in the score group i , Z_i .

Thus the probability function for person i to give the right response to item j is

$$P_{ij} = \frac{Z_i E_j}{1 + Z_i E_j} \qquad (1)$$

Among all models mentioned abovethe Rasch Model is the simplest one. The Rasch Model is described as a strong model(Lord,1967), but it simplifies the procedures for parameters estimation. Though the guessing factor is neglected in the model, it does not play an important rdle in test models. The guessing factor is a constant adding to the probability function, it can only shift the item characteristic curve by a constant equal to c_g, but cannot change the shape of the characteristic curve(Lord, 1967).

B.Wright(1969) who introduced a detail procedure for item analysis, based on the Rasch Model, claimed that the Rasch Model could lead to very high objectivity in psychological measurement. The calibration of item easiness was independent of the sample. The measurement of person ability was independent of subset of items used .

The procedure is called Sample Free Item Analysis (Wright,1969). It starts with the assumption that the probability of person i getting right answer on item j is given by the formula 1 .

Set $b_i = \log Z_i$

$d_j = \log E_j$ in the formula 1,

$$\text{then} \quad P_{ij} = \frac{\exp(b_i + d_j)}{1 + \exp(b_i + d_j)}$$

Any subject in the score group i of score c(i), is assumed to have the same ability d_i.

Thus,

$$P_{ij} \doteq \frac{a_{ij}}{r(i)} \doteq \frac{\exp(b_i + d_j)}{1 + \exp(b_i + d_j)}$$

where a_{ij} is an element of the data matrix A, of dimension NxM, which is equal to the number of persons in the score group i who get item i correct.

N = number of score groups

M = number of items

r(i) is an element of the score group matrix R, of dimension NxM, which is equal to the number of persons in the score group i of score c(i).

(Wright,1969,p.27).

$$\frac{r(i) - a_{ij}}{r(i)} \doteq 1 - P_{ij} = \frac{1}{1 + \exp(b_i + d_j)}$$

$$\frac{a_{ij}}{r(i) - a_{ij}} \doteq \exp (b_i + d_j)$$

$$t_{ij} = \log \frac{a_{ij}}{r(i)-a_{ij}} \doteq b_i + d_j \qquad (2)$$

where t_{ij} is an element in the Matrix T, of dimension NxM.

From the matrix T, the initial estimate of b_i and d_j can be obtained by the following formulae,

$$d_j = \frac{1}{N} \sum_{i=1}^{N} t_{ij} - \frac{1}{N\,M} \sum_{i=1}^{N}\sum_{j=1}^{M} t_{ij}$$

$$b_i = \frac{1}{M} \sum_{j=1}^{M} t_{ij} - \frac{1}{N\,M} \sum_{i=1}^{N}\sum_{j=1}^{M} t_{ij}$$

assuming that $\frac{1}{M}\sum_{j=1}^{M} d_j = 0$

$$\frac{1}{N}\sum_{i=1}^{N} b_i = 0$$

(Wright,1969,pp.27-29).

There will be troble when A is tranformed to T by the Formula 2 ,

at $a_{ij} = 0$, an empty cell in A ,

and $r(i)=a_{ij}$, a full cell in A.

In order to overcome this troble, t_{ij} is modified as,

$$t_{ij} = \log((a_{ij}+w)/(\underline{r(i)} - a_{ij} + w))$$

where $w = r(i)/ N$.

It is reasonable to do this, because

$$t_{ij} = \mp \log(1+N) \tag{3}$$

when $a_{ij} = 0$ or

$r(i)= a_{ij}$,

which is the upper limit of the magnitude of any element in T. The estimate of b_i and d_j can be improved by replacing the empty cell and full cells by,

$$t_{ij}' = b_i^* + d_j^*$$

where b_i^* = initial estimate of b_i

d_j^* = initial estimate of d_j

The procedure to estimate b_i and d_j is repeated with the modified matrix T.

The final estimates of b_i and d_j are obtained by the Maximum Likelood Iteration Method. This method can improve the estimate of parameters , so that they can maximize the probability of obtaining the pattern of responses.

Thus, $\sum_{i=1}^{N} a_{ij} - \sum_{i=1}^{N} r(i) \exp(b_i+d_j)/(1+\exp(b_i+d_j)) \doteq 0$

$$c(i) - \sum_{j=1}^{M} \exp(b_i+d_j)/(1+\exp(b_i+d_j)) \doteq 0$$

where $\sum_{i=1}^{N} a_{ij}$ is the number of subjects who get item j correct,

$c(i)$ is the score of the score group i.

The Newton Raphson procedure is used for the iteration process. The estimates of b_i and d_j, which are already obtained from the modified matrix T, are taken as the starting values of the iteration procedure. Improved values of b_i and d_j are then obtained. These improved values of b_i and d_j are taken as the starting values again and the procedure are repeated, until the estimates of the parameters do not change appreciably (Wright,1969, pp.30-33).

The iteration procedure is illustrated as the following,

$$d_j^{n+1} = d_j^{n} - f(d_j^{n})/f'(d_j^{n}) \qquad (4)$$

where $f(d_j^{n}) = \sum_{i=1}^{N} a_{ij} - \sum_{i=1}^{N} r(i) \exp(b_i^{n}+d_j^{n})/(1+\exp(b_i^{n}+d_j^{n}))$

$$f'(d_j^{n}) = -\sum_{i=1}^{N} r(i) \exp(b_i^{n}+d_j^{n})/(1+\exp(b_i^{n}+d_j^{n}))^2$$

$$b_i^{n+1} = b_i^{n} - g(b_i^{n})/g'(b_i^{n}) \qquad (5)$$

where $g(b_i^{n}) = r(i) - \sum_{j=1}^{M} \exp(b_i^{n} + d_j^{n})/(1+\exp(b_i^{n}+d_j^{n}))$

$$g'(b_i^{n}) = -\sum_{j=1}^{M} \exp(b_i^{n}+d_j^{n}) / (1+\exp(b_i^{n}+d_j^{n}))^2$$

in which, b_i^n and d_j^n are the starting values for ability in the score group i and easiness for item j respectively, in the iteration process, b_i^{n+1} and d_j^{n+1} are the improved values (Wright,1969,pp.34-36).

In the stage of item calibration, the starting values for ability and easiness are taken from Log method. A set of improved values of ability is obtained from the Formula 5. These improved values of ability are then put in the Formula 4 again to obtained another set of improved estimates of item easiness. The procedure is repeated to obtain better estimates both for ability and easiness.

However the solution of the parameters has an indeterminacy, for it can only be determined up to an unspecified constant adding to it (Rasch,1966,pp.50-53). If b_i and d_j are the solutions then it can be proved that

$$b_i' = b_i + k \quad \text{and}$$
$$d_j' = d_j - k$$

where k is any real number,

are also the solutions. The proof is illustrated by the following equations.

$$\exp(b_i' + d_j')/(1+\exp(b_i'+d_j'))$$

$$= \exp(b_i + k + d_j - k)/(1+\exp(b_i+k+d_j-k))$$

$$= \exp(b_i+d_j)/(1+\exp(b_i+d_j))$$

$$= P_{ij}$$

Wright and Panchapakasen(1969) also applied the Binomial Model to estimate the standard error of measurement(SEM) for bi and d_j respectively.

$$\mathrm{SEM}(b_i) = \frac{1}{M^2} \sum_{j=1}^{M} \left(1/(r(i)\, P_{ij}(1-P_{ij})) \right) \quad (6)$$

$$\mathrm{SEM}(d_j) = \frac{1}{N^2} \sum_{i=1}^{N} \left(1/(r(i)\, P_{ij}\, (1-P_{ij})) \right) \quad (7)$$

The SEM in the Formula 6 and 7 are over estimated . The actual value is less than that given by these two formulae (Wright,1969,pp.28-29).

When ability is measured, those score groups with more respondents in them, will have smaller errors of measurement. Also in the item calibration, the error will be minimal when

$$P_{ij}\,(1-P_{ij}) = \text{Maximum}$$

that is when $P_{ij} = 0.5$.

Thus the item with equal probability of pass and fail will have the minimum error of measurement. Although the Model is claimed to be sample free, the SEM is sample specific (Whitelt et al,1974,pp.170-171).

Another weakness of the model, which is often attacked, is the strong assumption of the model. The assumptions and implications other than the probability function of the model, are list as the following,

(1) ability has only one dimension ,
(2) minimal guessing of all items,

(3) subjects and items are locally independent, that is the response to an item from any subject does not affect the response of any other subject, and a subject's response to preceding items does not affect his response to later items(Whitely,1974),

(4) uniform discrimination of all items,

(5) all subjects who get the same score in the same test are assumed to have the same ability.

Because there are too many criteria that must be satified, the following questions concerning the model are often asked.

(1) Which kind of items can satisfy the assumptions?

(2) How can the goodness of fit of the model be tested?

Whitely(1974) suggested that those items fitting the model, must be culture free, have more choices of response and high reliability. Anderson(1968) and Wright(1969) had introduced two different approaches to test the goodness of fit of the model.

(I) Anderson's approach(t test)

The matrix T has an element t_{ij}.

$$t_{ij} = \log(P_{ij}/(1-P_{ij}))$$

$$t_{ij} = b_i + d_j \tag{8}$$

The vector of row mean of the Matrix T has an element u_i.

$$u_i = \frac{1}{M} \sum_{j=1}^{M} t_{ij}$$

$u_i = b_i + d.$

$$\text{where } d. = \frac{1}{M} \sum_{j=1}^{M} d_j$$

The difference between t_{ij} and u_i is a constant for any i, thus when only item j is considered and is assumed to fit the model, that is when Formula 8 is valid, plotting t_{ij} against u_i, will give a straight line with slope equal 1 (Anderson,1968). This method was also employed by G. Rasch to test the fit. The slope L of the straight line is given by the following formul, assuming that t_{ij} and u_i have linear relation

$$\text{Slope} = L = \frac{\sum_{i=1}^{N} t_{ij}\, u_i - \frac{1}{N}\left(\sum_{i=1}^{N} t_{ij} \sum_{i=1}^{N} u_i \right)}{\sum_{i=1}^{N} (t_{ij})^2 - \frac{1}{N}\left(\sum_{i=1}^{N} t_{ij} \right)^2} \qquad \text{....................(9)}$$

(Draper, et al, 1966, pp. 12-13).

To test whether the slope L equals 1, the t value is calculated by the Formula 10.

$$t(N-2,\ 1-\frac{1}{2}\alpha) = \frac{L - 1.0}{\text{EST.S.E.}(L)} \qquad (10)$$

where EST.S.E.(L) is the estimated standard error of L,

$$\text{EST.S.E.}(L) = S / \left(\sum_{i=1}^{N} (t_{ij} - t._j)^2 \right)^{1/2}$$

$$\text{where } S = (SS/(N-2))^{1/2}$$

$$SS = L \left(\sum_{i=1}^{N} t_{ij}\, u_i - \frac{1}{N}\left(\sum_{i=1}^{N} t_{ij} \sum_{i=1}^{N} u_i \right) \right) - \left(\sum_{i=1}^{N} u_i^{\,2} - \frac{1}{N}\left(\sum_{i=1}^{N} u_i \right)^2 \right)$$

The t value of degree of freedom N-2 can test the departure from unit slope for each item. If the t value is not significant, the item is considered to fit the Model.

(II) Wright's approach (x^2)

B.Wright's approach assumes binomial distribution. This approach can tell how much the estimated parameters fit the model and the response data.

Let y_{ij} be an element in the Matrix Y.

$$y_{ij} = \frac{a_{ij} - E(a_{ij})}{(\ Var(a_{ij}))^{1/2}}$$

where $E(a_{ij})$ is the expectation value of a_{ij} ,

$$E(a_{ij}) = r(i)\ P_{ij}$$
$$= r(i)\ exp(b_i+d_j)\ /\ (1+exp(b_i+d_j))$$

Assuming binomial distribution , the variance of a_{ij} is given by the following formula.

$$\begin{aligned} Var(a_{ij}) &= nPq \\ &= n\ P_{ij}\ (1-P_{ij}) \\ &= r(i)\ exp(b_i+d_j)/(1+exp(b_i+d_j))^2 \end{aligned}$$

Y is a NxM matrix. The chi square statistics for each item is computed to test the goodness of fit of the estimation of parameters , by the following formula.

$$x^2 = \sum_{i=1}^{N} (y_{ij})^2 \qquad (11)$$

degree of freedom = N-1

No significance in the chi square statistics for item j implies that this item can achieve good estimation of parameters and is considered as fitting the model (Wright, 1969, pp.44-45).

There are also different kinds of statistical approach to test the objectivity of the model. The following statistics are often be used.

(I) Pearson Moment Product Correlation Coefficient

Significantly high correlation coefficient of person ability calibrated from two different set of items, or of item easiness calibrated from two different sample, indicates that the model is very objective.

(II) T test (correlated and noncorrelated)

No significance in t ratio, supports the claim that that the parameters of the model are sample free and item free.

(III)Standardized Difference Scores

These scores can be interpreted as the Z scores (Whitely,1974). It is equal to the difference of two scores obtained from two set of tests, divided by the combined measurement error associated with two sets of tests.

$$D_{12} = \frac{x_1 - x_2}{(SEM^2_{x_1} + SEM^2_{x_2})^{1/2}}$$

where D_{12} = standardized difference score

x_1 = ability calibrated from the first test

x_2 = ability calibrated from the second test

$SEM^2_{x_1}$ = square of the standard error of measurement associated with x_1

$SEM^2_{x_2}$ = square of the standard error of measurement associated with x_2

These scores were employed in Whitely's study to investigate the equivalence between two tests. If two tests are statistically equivalent, the standardized difference scores(SDS) will be normally distributed with a mean of zero and standard deviation of one(Whitely, 1974). Wright(1967) had also used the SDS to study the item free property of the model .

Among these three kinds of statistics, the SDS is currently used and regarded as a more limited definition of equivalent measures(Wright, 1967. Whitely,1974).

Although invariance of parameters Of the Rasch Model was reported by Anderson(1968), the statistics he used was the correlation coefficient, which could not completely support the claim of invariance.

Whitely(1974) reported the the equivalence of ability in the Rasch Model, calibrated from two different item subsets, failed if they were tested by the standardized difference scores. But unfortunately, only 30 percent of items in his study fitted the model, and the invariance of item easiness under different samples was not studied.

This study purports to examine the equivalence of item parameters calibrated from two different samples(item easiness) and test subsets(person ability). The goodness of fit of the model of each item will be tested by the chi square. Only these items that fit the model will be taken in the study. The equivalence of the model parameters will be tested by the correlation coefficient, t test (correlated and noncorrelated), and the standardized difference scores.

Related Literature

The Rasch Model intriduced in 1966, had been proved theoretically that subject parameters(person ability) could be evaluated independently of item parameters and without regard to parameters of other subgroups(Rasch,1966). Since item easiness and person ability are symmetric in the model, similar result will also be obtained for item easiness. This property was called specific objectivity by Rasch.

After that, many researchers studied the objectivity and robustness of the Rasch model empirically.

(I) Evidence of the objectivity of the Rasch Model

Anderson(1968) used a 45-item Spiral Omnibus intelligence test in his study. The test was administered to two different samples. The following results were reported.

1. item difficulty indices were independent of the sample on which they were based,

2. ability indices were independent of the sample on which they were based.

Wright(1968) studied the responses of 976 beginning Law students to 48 reading comprehension items of the Law School Admission Test and obtained the following result.

1. item calibration was sample free, independent of the particular persons used for the calibration.

2. person ability could be estimated from the Easy test or from the Hard test, and the results showed no difference.

Although quit a lot of research supports the claim of objectivity of the Rasch Model, Whitely(1974) reported that the equivalence of person ability calibration based on two different item subsets failed under more rigorous definition of equivalent measurement.

(II) Robustness of the model

Robustness of the model was also reported in several studies.

In Anderson's study(1968), 20.4 percent of items failed to fit the model at 0.05 level, but still, evidence of objectivity of the model was reported. Wright(1969) claimed that the model was quite robust with respect to departure from the assumption of uniform discrimination and guessing factor.

A different finding was obtained by Whitely et al (1974). He rejected the hypothesis that there was no significant difference between person ability calibrated on the easy test and the hard test. The test items that he used did not fit the model completely. Thus it can be believed that the robustness of the model is questionable.

(III) Meaning of freeness of the Rasch Model

The definition of freeness of the Rasch model among researches on the model, is not very clear. Anderson and Wright both reported the person free property of item easiness calibration. The statistics that they used were different. The Table 1 illustrates the difference between them.

The Table 1 also shows that the person free of item calibration is not defined very clearly by researchers.

However the meaning of item free person ability measurement is quite consistent among researchers. The statistics used by Wright(1967) and Whitely(1974) to test the equivalence of ability measurement based on two different item subsets, are the standardized difference scores. Zero mean and unit standard deviation of SDS indicates the statistical equivalent of two item subsets. In Whitely's study, the conventional statistics , correlated t test and correlation coefficient were also reported.

TABLE 1

Summary of the meanings of freeness

Author	Samples	Statistics for test of person free	test of fit to model	did all items fit ?
J. Anderson (1966)	C.M.F./ R.A.N.[a]	very high r reported[b]	t test of unit slope	yes
	C.M.F./ R.A.N.	very high r reported	t test of unit slope	no
B.D. Wright (1968)	high score/ low score group	no, only illustrated by graphs	chi square	no

a C.M.F.= applicants of Citizen Military Forces
R.A.N.= applicants of Royal Australian Navy

b r = Pearson Moment Product correlation Coefficient

Definitions

Some common terms in the study are defined operationally as follows:

1. Person ability and item easiness are defined by the Rasch model, so that the right response to an item j of easiness E_j, by a subject Z_i is given by

$$P_{ij} = \frac{E_j Z_i}{1 + E_j Z_i}$$

2. Low score group is defined as those subjects with scores less than or equal to 13 in the Test C. The maximum score is 30.

3. High score group is defined as those subjects with scores greater than or equal to 14 in the Test C. The maximum score is 30.

4. Hard test is defined as a set of test items with low item easiness, so that the item easiness of any item in the Hard test is less than that of any item in the Easy test.

5. Easy test is defined as a set of test items with high item easiness, so that the item easiness of any item in the Easy test is greater than that of any item in the Hard test.

6. Item fitting the model is defined as the item of no significant value at 0.05 in the chi square statistics defined by the Wright's procedure.

7. Test A is the test used in the pilot test. It is a Physics achievement test with 100 items.

8. Test B is the test used in the experiment. Items in the test B are selected from the Test A following the criteria of similar point biserial correlation coefficients among items(ranging from0.15 to 0.5). There are totally 47 items in the test.

9. Test C is defined as the test with 30 items selected from the test B,following the criteria of fitting the Rasch model .

Hypotheses

In order to test the independence of item subsets in person ability measurement, and the independence of samples in the item easiness calibration, the following hypotheses were formulated:

1. As calibrated by the Rasch model, there is no significant difference between a student's ability as measured by the Hard test and the Easy test.

2. As calibrated by the Rasch model, there is no significant difference in item easiness of the item in Test C, between the High score group and Low score group.

CHAPTER TWO

METHODOLOGY

Research Design

In order to test hypothesis 1 , ability of the sample was calibrated from the Hard test, then from the Easy test, by the Rasch-Wright Procedure.

For testing hypothesis 2 , the easiness of those items fitting the model was calibrated first from the Low score group, then from the High score group, following the Rasch-Wright Procedure.

The design is illustrated in Figure 1. The significance of difference was test by the Pearson Moment Product Correlation Coefficient, t test(correlated t test for I, noncorrelated t test for II), and the standardized difference score(for I only), at the same time.

I

Hard test	Easy test
person ability	person ability

II

Low score group	High score group
item easiness	item easiness

FIG. 1 Illustration of research design

Sampling

Subjects were selected from among Form 3 and Form 4 secondary school students in Hong Kong. They belonged to six different schools, including one government school, three subsidized schools and two private schools.

In the Hong Kong setting, it is general practice that most of the high ability students are allocated to government and subsidized schools, while low ability students are allocated to private schools. Thus , the distribution of the abilities can be assumed to cover a large range. The total number of subjects in the sample was 464. Those extreme score groups(extremely high or extremely low) can have enough respondents to achieve a stable estimation of probability.

The instrument in this study was an achievement test in Physics. In view of the fact that the teaching syllabuses in the sample schools were not at the same level, the selection of Form3 or Form 4 students for the experiment was made according to the teaching syllabus in each particular school.

Instruments

The following is a list of instruments employed in the study.

1. Test A(Physics achievement test with 100 items)

Items in the Test A were constructed either by the researcher or modified from tests in published test books, on the following topics:

a. meaning of latent heat,
b. meaning of specific heat capacity and heat capacity,
c. transfer of heat energy when there is a change of temperature,
d. transfer of heat energy when there is a change of state,
e. conservation of heat energy during the process of heat transfer.

All items were carefully constructed so that high validity could be built into the test and so that there was a high homogenaity of content in the test.

2. Test B(with 47 items)

Items in the Test B were selected from the 100-item Test A after the pilot test was completed. They were selected according to the following criteria:

a. medium difficulty level,
b. similar values of point biserial correlation coefficients(ranging from 0.15 to 0.5).

In order to minimize the measurement error, the very difficult items were deleted. Those items with similar point biserial correlation coefficients could be assumed to have uniform discrimination power.

For item analyses, which followed the Rasch-Wright Procedure, a computer programme was developed by the researcher and executed by the ICL 1900 Computer in the Computer Center of the Chinese University of Hong Kong. The input to the programme was the response matrix on a test. It could go through the Rasch-Wrigh Procedure and produce the item easiness, ability of score group, calibration of person ability of all subjects in the sample and the measurement error for item easiness and person ability.

A simplified flowchart of the programme is shown in the Figure 2. In the iteration process, no criterian of termination was built into the programme. Usually the process would be converged for at most 10 iterations, according to the researcher's experience. Therefore the maximum number of iterations in this programme was set to 20.

In the stage of person ability calibration (from the Hard test and from the Easy test), the item easiness was assumed known. When ability was calibrated from the Hard test, the item easiness of all items in the Hard test, which were already calibrated, were input in the iteration stage of the programme. Similar procedures were taken when person abilities were calibrated from the Easy test.

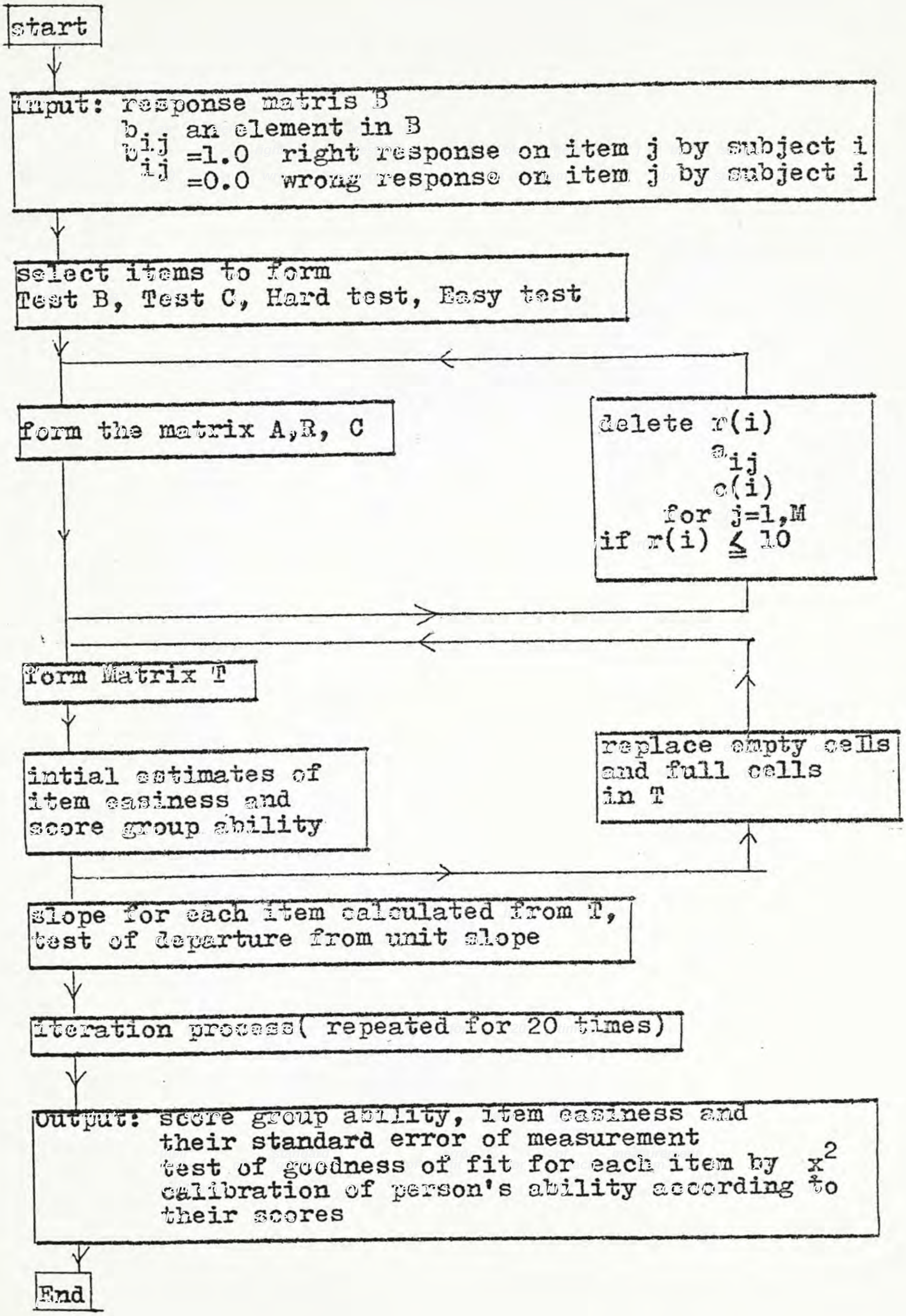


FIG. 2 A simplified flowchart of the programme

The programme deleted those score groups containing subjects fewer than 10, so as to minimize the measurement errors as well as maximize the convergence of the iteration process.

Experimental Procedure

Before the experiment was actually carried out, a pilot test was conducted to test the instrument. The Test A with 100 items was administered to 120 Form 4 students in a school not used for the actual experiment. The average time required to complete each item was estimated. The results from the pilot test was analysed. The test reliability(Hoyte's coefficient alpha), point biserial correlation coefficient and difficulty level of each item were calculated.

A total of 47 items, which satisfied the criteria set up in this study, were selected from the 100-item Test A. These items made up the final instrument used for the study. It was named as the Test B. It was then administered to the sample. The subjects were told to mark their answers on the answer sheets provided. In order to minimize the guessing factor, enough time was allowed for the students. In addition, teacher administering the test was asked to announce to the students that they should not attempt to guess the answers.

All completed sheets were corrected by hand. Each item was coded, 1 for right response, 0 for wrong response. The response matrix of all items in the from all subjects was put in the storage file in the ICL 1900 computer system for analysis.

Method of analysis

The responses for the items of Test B were calibrated by the Rasch-Wright Procedure to the test the goodness of fit for each item. The procedure was carried out by the ICL 1900 computer.

Those items fitting the model(in terms of chi square statistics) were selected. There was a total of 30 items which were considered best fitting to the model . The item easiness of each of this 30 items in the Test C, was calibrated again from responseof (1) the entire sample (2) the high score group and (3) the low score group. The Test C was then divided into two subsets of equal number of items, the Hard test and the Easy test, according to their item easiness.

Person ability of all subjects in the sample were calibrated by the Hard test, and then the Easy test separately. During the process of ability calibration, the item easiness was assumed known. Because the Hard test and the Easy test were subsets of the Test C, their item easinesses were obtained from the calibration process of the Test C. The Figure 3 illustrates the calibration procedure.

Two sets of item easiness calibrated from different groups(the high score and low score groups) were compared and analysed by the Pearson Moment Product Correlation Coefficient, t test and standardized difference scores. Similar methods of analyses were employed for two sets of person ability measured by different test subsets.

Sample \ Test subsets	Test B 47-item	Test C 30-items	Hard test 15-item	Easy test 15-item
entire sample	X[a]	X	X	X
High score group		X		
Low score group		X		

a X indicates the cell where the calibration process was performed

FIG. 3 The calibration process

CHAPTER THREE

RESULTS AND DISCUSSIONS

The results of the study were reported in terms of three aspects, the selection of items fitting the model, the invariance of item easiness and the invariance of person ability.

1. The selection of items of good fit

The responses on the Test B from all subjects were analysed by the computer. The output of the programme gave the following information :

I. item characteristics
   a. difficulty level,
   b. point biserial correlation coefficients,
   c. the index of the goodness of fit for each item as described by the chi square and the slope.

II. item easiness,

III. score group ability,

IV. calibration of ability for each person according to his score obtained in the test.

The Table 2 shows the item characteristics of each item in the Test B. There were 30 items in the 47-item Test B

TABLE 2

Item charateristics and goodness of fit for each item in the Test B as calibrated from the entire sample

item no.	difficulty level	point biserial corr.	chi-square df=35	slope df=34
(1)	0.33190	0.19088	76.16152 *	0.86530
(2)	0.70905	0.20099	49.81142 *	0.92935
(3)	0.70474	0.26509	41.97001	0.90041
(4)	0.29310	0.32230	24.80852	0.80740**
(5)	0.59267	0.39422	34.56570	0.82762**
(6)	0.33621	0.34618	31.58543	0.79500**
(7)	0.49353	0.22625	60.72177 *	0.82062
(8)	0.61207	0.33426	34.92584	0.85805**
(9)	0.42457	0.32362	38.72675	0.80847 **
(10)	0.55819	0.28833	37.59115	0.86004
(11)	0.37931	0.36678	25.47307	0.83045 **
(12)	0.21336	0.20525	38.47675	0.86473
(13)	0.62931	0.22834	46.22608 *	0.83464 **
(14)	0.91595	-0.01250	97.31745 *	0.66404 **
(15)	0.38793	0.40056	46.38358 *	0.77969 **
(16)	0.78448	0.05088	72.36423 *	0.80071 **
(17)	0.82112	0.32334	29.14213	0.84476 **
(18)	0.69397	0.30914	31.64345	0.83089 **
(19)	0.62284	0.34753	22.14523	0.84590 **
(20)	0.64224	0.36563	24.65827	0.86490 **
(21)	0.69181	0.26725	27.62610	0.89451
(22)	0.76293	0.21209	46.84000 *	0.88222
(23)	0.45259	0.42472	39.20115	0.80184 **
(24)	0.38578	0.44657	26.22708	0.82470 **
(25)	0.44612	0.42172	29.90885	0.87652
(26)	0.86853	-0.05437	117.46483 *	0.68294 **
(27)	0.42888	0.45743	26.32423	0.82842 **
(28)	0.64009	0.35242	27.53560	0.87479
(29)	0.32543	0.36352	31.86569	0.82661 **
(30)	0.47198	0.44740	24.18833	0.82486 **
(31)	0.61207	0.22116	56.31703 *	0.84326 **
(32)	0.78233	0.21395	69.77262 *	0.83606
(33)	0.62026	0.44389	55.81186 *	0.83691 **
(34)	0.43966	0.49230	39.11850	0.79532 **
(35)	0.54741	0.49853	32.90364	0.80817 **
(36)	0.70905	0.38597	31.21330	0.83950 **
(37)	0.49569	0.41855	56.20125 *	0.81461
(38)	0.41810	0.52951	38.80095	0.77250 **
(39)	0.60991	0.49055	33.68528	0.77570 **
(40)	0.55388	0.42703	34.76015	0.81657 **
(41)	0.32112	0.49792	44.37643 *	0.75074 **
(42)	0.51940	0.48169	33.42593	0.81387 **
(43)	0.73276	0.00843	94.16903 *	0.79033
(44)	0.53448	0.44629	22.05732	0.84058 **
(45)	0.74138	0.18814	49.50031 *	0.82078 **
(46)	0.72414	0.37881	27.97216	0.83390 **
(47)	0.92241	-0.01084	81.98396 *	0.69241 **

* significant in chi square at 0.05

** significant in departure from unit slope at 0.05

which were not found to be statistically significant by the chi square method. Thus they were used to make up another test, the Test C.

As illustrated by the Figure 4, the point biserial correlation coefficients for most items in the Test B, took values from 0.2 to 0.5. It seemed that the discrimination indice among them did not differ very much, so that we could state that the items in the Test C had satisfied the assumption of uniform discrimination.

Figure 5 shows another approach to test the goodness of fit for the items. As it can be seen from the graph, most of the items have slope in the range from 0.8 to 0.9. thus in general, most items fitted good in the model.

2. The invariance of item easiness calibrated from different groups

Item easiness of the Test C were calibrated from the entire sample, the high score group and the low score group. The Tables 3,4, and 5 show the item characteristics of the Test C calibrated in different groups. As it can be seen from the tables, large number of the items in the Test C can fit the model in terms of chi square statistics. But the t test for unit slopeshows that most items depart significantly from unit slope. It seemed that the items with significant chi square also had larger departure from the unit slope. As shown in the Table 3,4, and 5, the

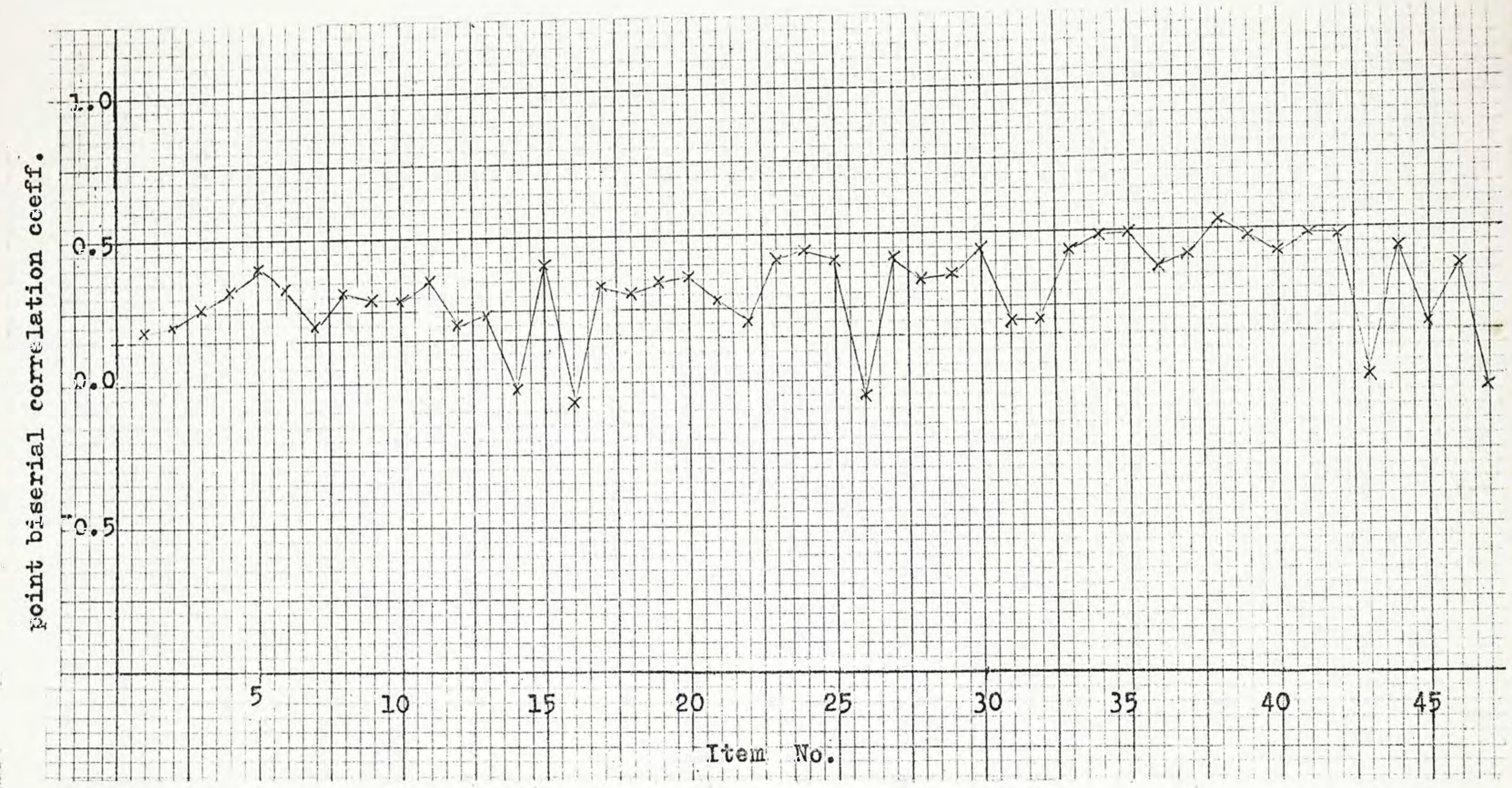


FIG. 4 Point biserial correlation coefficients for each item in the Test B as calibrated from the entire sample

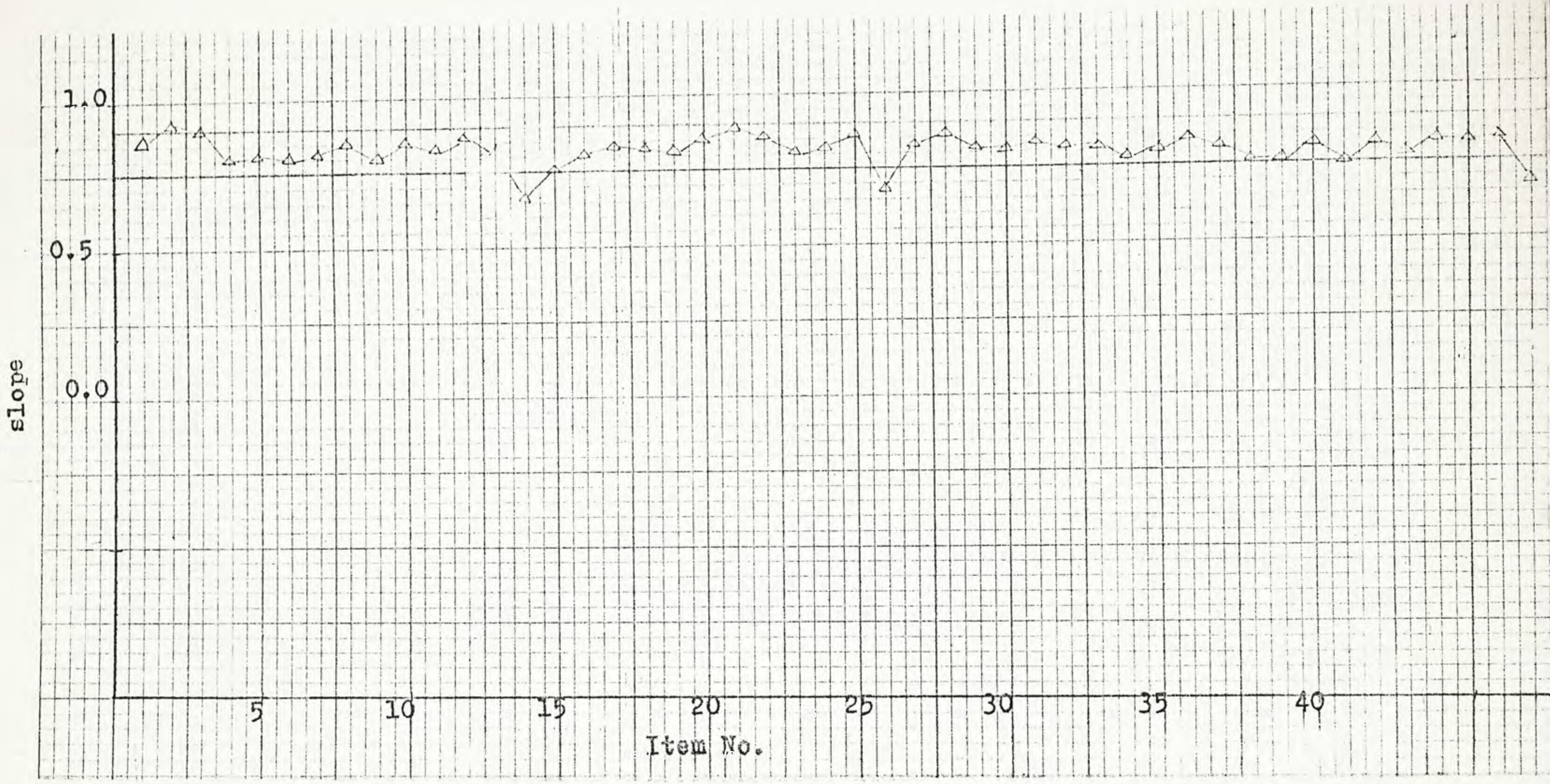


FIG. 5 Slope calculated for each item in the Test B as calibrated from the entire sample

TABLE 3

Item characteristics and the goodness of fit for each item in the Test C as calibrated from the entire sample

item No.	difficulty level	point biserial corr.	chi-square df= 19	slope df=18
(1)	0.70474	0.25691	21.41161	0.99305
(2)	0.29310	0.36490	13.81233	0.86716
(3)	0.59267	0.42026	17.07592	0.85193
(4)	0.33621	0.38106	15.80180	0.88259
(5)	0.61207	0.35596	18.25840	0.93313
(6)	0.42457	0.35775	13.72862	1.08015
(7)	0.55819	0.31896	23.17893	1.17201
(8)	0.37931	0.41592	12.79935	0.74746 **
(9)	0.21336	0.23100	27.04062	0.91900
(10)	0.82112	0.33920	12.96510	0.81790
(11)	0.69397	0.31364	46.92620*	0.61081 **
(12)	0.62284	0.36409	40.54741*	0.70409 **
(13)	0.64224	0.39149	19.06627	0.87581
(14)	0.69181	0.31448	31.83069*	0.88268
(15)	0.45259	0.44463	19.57032	0.75991 **
(16)	0.38578	0.46280	18.99052	0.70352 **
(17)	0.44612	0.43344	9.19450	0.77771 **
(18)	0.42888	0.47423	23.92085	0.71884 **
(19)	0.64009	0.36660	15.71410	0.97149
(20)	0.32543	0.36179	15.05693	0.80111 **
(21)	0.47198	0.45749	18.33378	0.75361 **
(22)	0.43966	0.48497	24.62382	0.66486 **
(23)	0.54741	0.49366	14.93631	0.68128 **
(24)	0.70905	0.40277	24.93752	0.84124
(25)	0.41810	0.52900	25.26431	0.56253 **
(26)	0.60991	0.51144	24.27754	0.66868 **
(27)	0.55388	0.38695	19.84907	0.83603
(28)	0.51940	0.47553	16.52374	0.72564 **
(29)	0.53448	0.46086	20.46292	0.64737 **
(30)	0.72414	0.40465	23.81055	0.67189 **

* significant in chi square at 0.05

** significant in departure from unit slope at 0.05

TABLE 4

Item characteristics and goodness of fit for each item in the Test C as calibrated from the high score group

item No.	difficulty level	point biserial corr.	chi-square df= 10	slope df= 9
(1)	0.60280	0.14385	14.82518	0.53713
(2)	0.16355	0.14992	9.14071	0.50439
(3)	0.45327	0.26806	7.99746	0.58115 **
(4)	0.20561	0.12463	6.80312	0.66739
(5)	0.50467	0.13763	8.70689	0.72056
(6)	0.30841	0.19561	9.63231	0.53708
(7)	0.43458	0.04873	13.23605	0.76374
(8)	0.25234	0.30429	8.40372	0.57790 **
(9)	0.15421	0.13645	11.68260	0.32576
(10)	0.79907	0.16494	4.96215	0.47561 **
(11)	0.62617	0.13829	22.64307*	0.28325 **
(12)	0.56542	0.11701	25.23743*	0.30779 **
(13)	0.55140	0.20041	8.67692	0.55936 **
(14)	0.61215	0.24645	13.22613	0.47075 **
(15)	0.27103	0.18466	14.13043	0.33735 **
(16)	0.23832	0.28848	11.90843	0.52547 **
(17)	0.28972	0.21540	5.88752	0.52109 **
(18)	0.23364	0.22881	14.04091	0.40441 **
(19)	0.54673	0.22704	10.95616	0.59704
(20)	0.20561	0.24253	6.43450	0.54905 **
(21)	0.30841	0.09244	9.67752	0.54815
(22)	0.26636	0.27533	12.19526	0.53444 **
(23)	0.38785	0.26274	9.19449	0.47854 **
(24)	0.63084	0.34593	16.78919	0.46654 **
(25)	0.22897	0.35995	14.63344	0.46866 **
(26)	0.44860	0.26871	10.17009	0.48510 **
(27)	0.45794	0.22456	14.12173	0.44865 **
(28)	0.37850	0.23768	12.28794	0.52610 **
(29)	0.37850	0.22074	10.99968	0.48690 **
(30)	0.61682	0.13844	16.33611	0.44947 **

* significant in chi square at 0.05

** significant in departure from unit slope at 0.05

TABLE 5

Item characteristics and the goodness of fit for each item in the Test C as calibrated from the low score group

item No.	difficul -ty level	point biserial corr.	chi- square df= 8		slope df= 7	
(1)	0.82143	0.00535	6.40632		0.42143	
(2)	0.44643	0.19330	4.90695		0.75421	
(3)	0.77232	0.14530	9.07005		0.40096	
(4)	0.49554	0.25271	9.12533		0.44771	**
(5)	0.75893	0.20568	7.86382		0.58697	
(6)	0.57143	0.21021	1.87935		1.12581	
(7)	0.70982	0.17711	8.25246		0.46439	
(8)	0.54018	0.24643	4.01053		0.58485	
(9)	0.28125	0.23206	4.25305		0.59083	
(10)	0.90179	0.19771	3.64064		0.66711	
(11)	0.81250	0.11080	19.08102	*	0.12954	**
(12)	0.74554	0.28142	8.22048		0.46137	**
(13)	0.79018	0.17733	9.11631		0.45159	**
(14)	0.79911	0.16789	11.73786		0.37563	**
(15)	0.66071	0.25265	5.38934		0.50463	**
(16)	0.57143	0.28653	7.00795		0.52630	**
(17)	0.63393	0.26965	3.16405		0.55351	**
(18)	0.64732	0.28166	7.71514		0.61798	
(19)	0.77679	0.18393	2.97467		0.60343	**
(20)	0.46875	0.20416	8.38460		0.44084	**
(21)	0.67857	0.32233	8.56288		0.42367	**
(22)	0.65179	0.26462	10.83822		0.45262	**
(23)	0.75893	0.24453	4.20145		0.49644	**
(24)	0.83929	0.15969	7.02091		0.49694	**
(25)	0.64286	0.31439	8.09901		0.34669	**
(26)	0.83036	0.18955	10.66573		0.49656	**
(27)	0.69643	0.26978	2.95558		0.60163	**
(28)	0.71429	0.28951	4.16943		0.51582	**
(29)	0.73661	0.24147	9.17963		0.31962	**
(30)	0.87946	0.24182	7.36056		0.44188	**

* significant in chi square at 0.05

** significant in departure from unit slope at 0.05

item No.11 of the Test C , did not fit the model in the sample, the high score group and the low score group. Its slope was the minimum among the items.

Figure 6 shows the point biserial correaltion coefficients of the items in the Test C when it was calibrated from the entire sample, the high score group and the low score group. Most of them lies in the range from 0.15 to 0.45. The criterion of uniform discrimination among items seems to be satisfied. The slope for each item in the Test C is illustrated in the Figure 7. It shows that the Test C when it was calibrated from the entire sample, departed less from the unit slope. Testing the goodness of fit to the model is sample biased . A test fitting the model in this sample may fail to fit in another sample.

Item easiness of the Test C calibrated from the high score group and the low score group is shown in the Table 6. The Figure 8 shows their values in a graph. As shown by the graph, they were no completely identical. The pattern of the curves suggested that two sets of item easiness differed by a fairly constant value. Most of the item easiness from the low score group could be made approximately equal to that from the high score group, if a value(approximately = 1.5) was added to each individual item.

Table 7 gives a statistical comparision of the item easiness of the items in the Test C, calibrated from the high score group and the low score group. The correlation coefficient between two sets of item easiness,

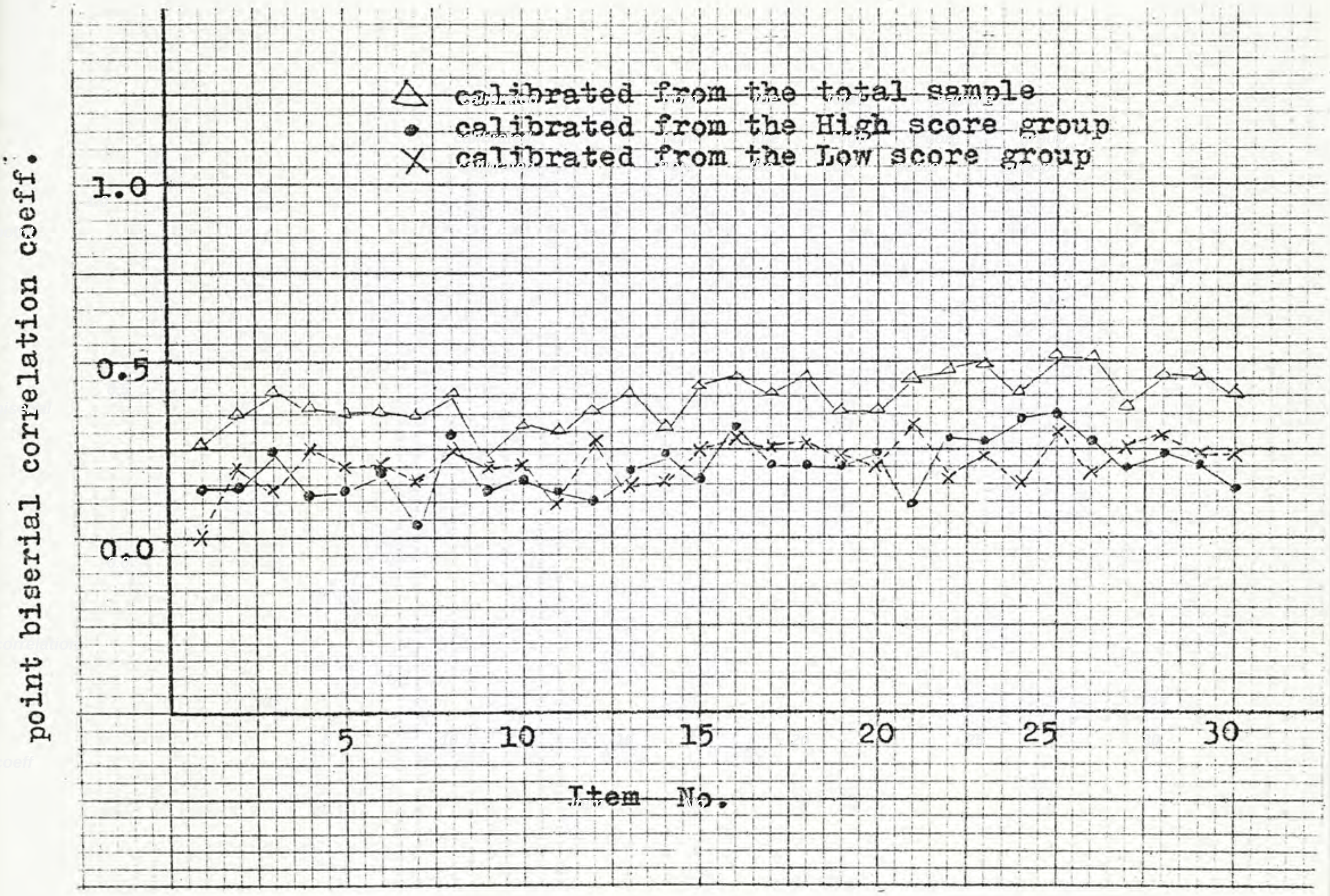
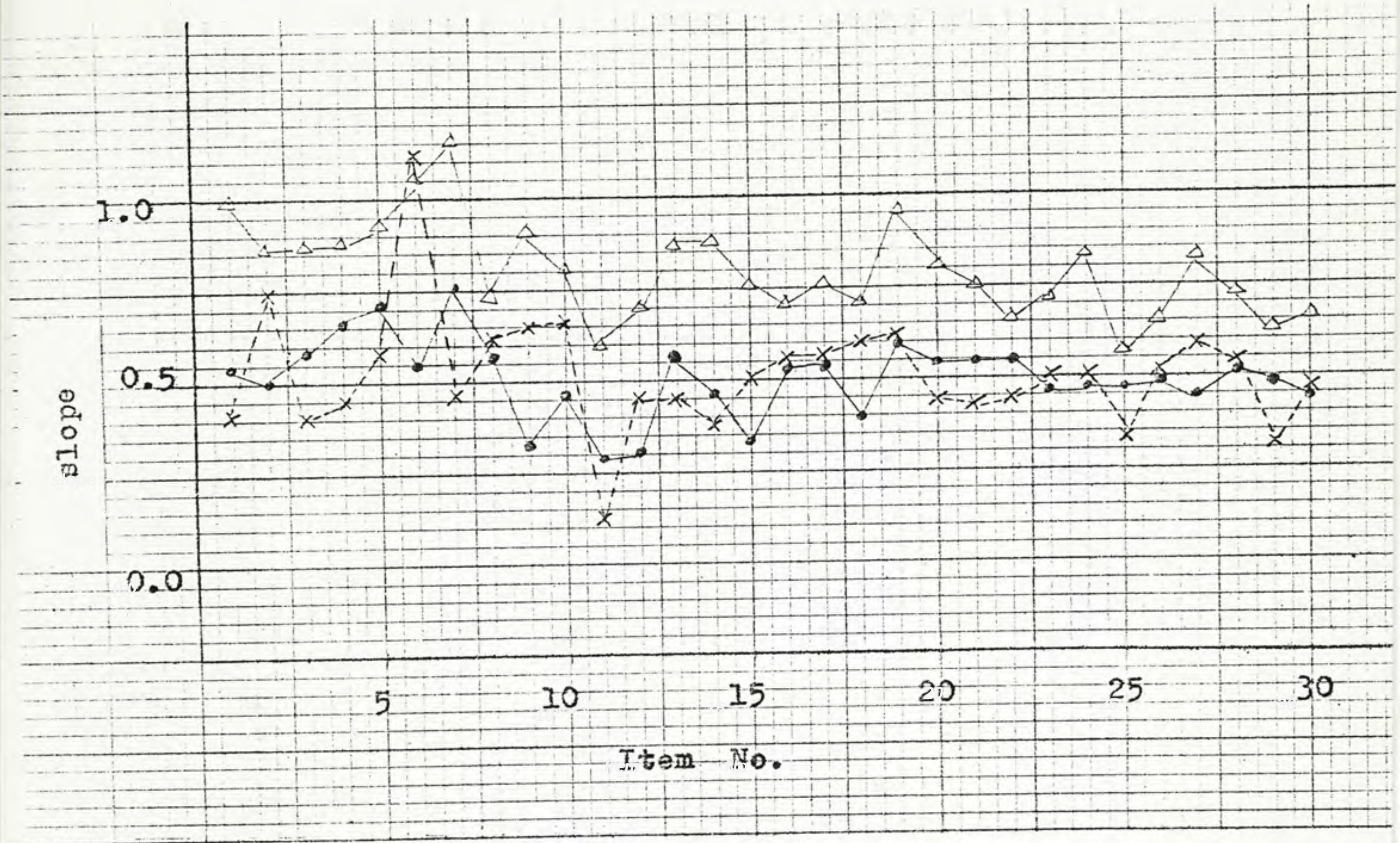


FIG. 6 Point biserial correlation coefficients for the Test C as calibrated in different groups



△ calibrated from the total sample

• calibrated from the High score group

X calibrated from the Low score group

FIG. 7 Slope calculated for each item in the Test C as calibrated from fifferent groups

TABLE 6

Item easiness of the Test C as calibrated from the entire sample, the low socre group and the high score group

Item No.	From High score group	From Low score group	From total sample
(1)	-0.29225	-1.80147	-0.98795
(2)	1.88426	0.14180	1.07448
(3)	0.37970	-1.31516	-0.38690
(4)	1.60318	-0.04210	0.85386
(5)	0.17010	-1.23624	-0.47604
(6)	1.00730	-0.36624	0.40132
(7)	0.43738	-0.96748	-0.20103
(8)	1.36023	-0.22418	0.64538
(9)	1.91871	0.93515	1.57681
(10)	-1.25345	-2.34820	-1.82318
(11)	-0.31212	-1.60322	-0.92454
(12)	-0.03942	-1.16001	-0.56662
(13)	-0.00116	-1.45375	-0.67061
(14)	-0.33207	-1.48269	-0.88710
(15)	1.18915	-0.72151	0.30365
(16)	1.43791	-0.38665	0.58906
(17)	1.11947	-0.65713	0.30365
(18)	1.38580	-0.72151	0.39042
(19)	-0.02028	-1.36945	-0.64729
(20)	1.60318	0.01386	0.88966
(21)	1.05162	-0.80396	0.19601
(22)	1.28531	-0.72151	0.34694
(23)	0.69290	-1.26223	-0.19021
(24)	-0.33207	-1.76672	-1.00080
(25)	1.49142	-0.63587	0.47802
(26)	0.45669	-1.73271	-0.49854
(27)	0.39889	-0.92137	-0.20103
(28)	0.75369	-0.99081	-0.03957
(29)	0.75369	-1.11051	-0.09322
(30)	-0.27246	-2.11237	-1.07910

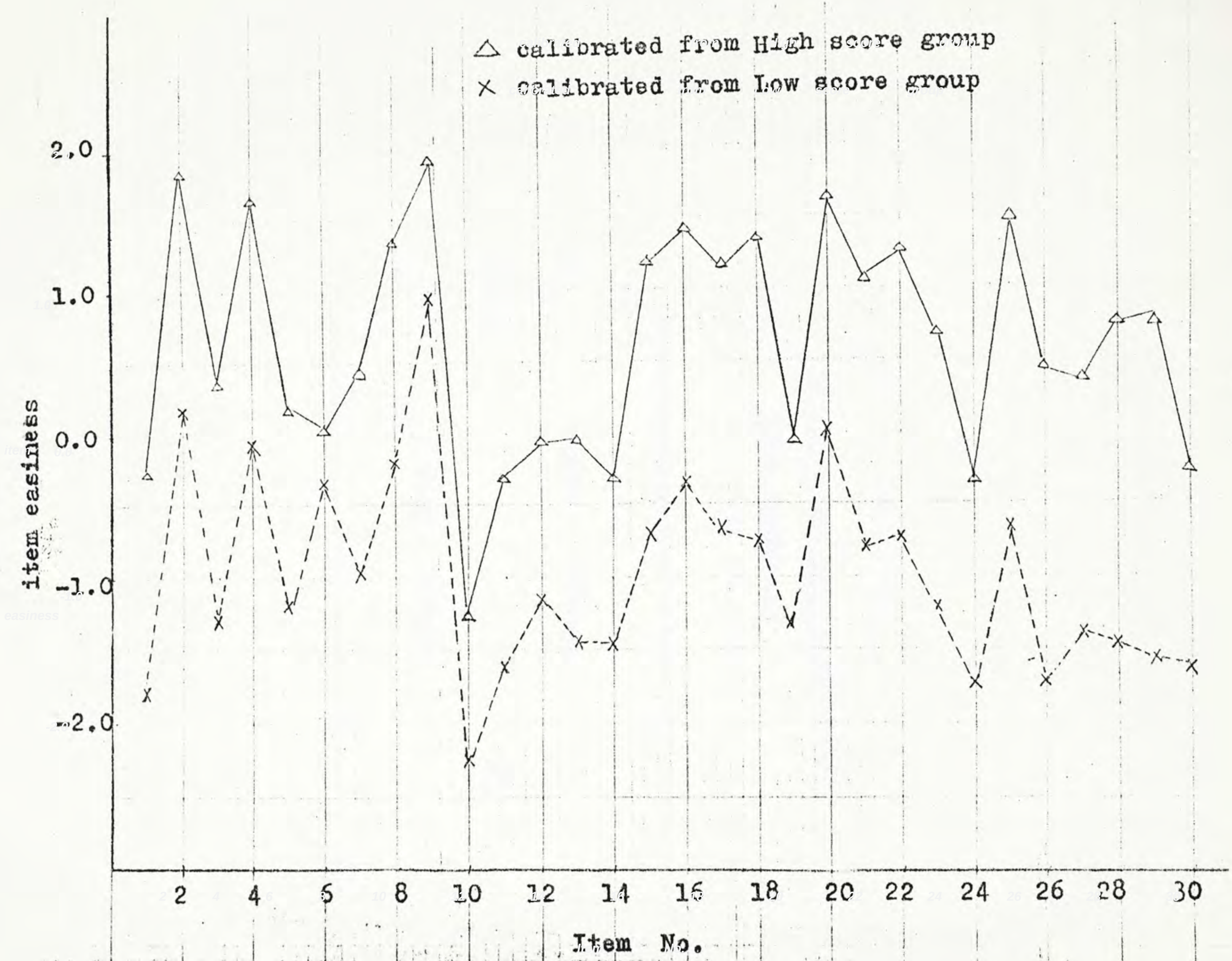


FIG. 8 Item easiness of the Test C as calibrated from different groups

TABLE 7

Comparis ons of item easiness of the Test C as calibrated from the high score group and the low score group

calibration sample	r	Item easiness $\overline{X}$	s	t	F_{max}
High score group		0.6175	0.8594		
v.s.	0.8974			7.726**	1.4387
Low score group		-0.9608	0.7165		

** $p < 0.05$

their mean $\bar{x}$, and standard deviations, t test for testing their mean difference, Hartley's test F_{max} for testing their variance difference were reported. There was a very high correlation between them, but not significantly in Hartley's test. This was quite consistent with the finding of Anderson(1966).

The t test shows significant differences. This is expected. As suggested by the graph in the Figure 8, they may differ by a constant value. Rasch (1966) had pointed out that any solution of the item parameters in the model could have a difference of any constant.

Other item parameters in the Rasch model, the score group ability of the Test C calibrated from the entire sample, the high score group and the low score group, are listed in the Table 8. The low score group gave the calibration of low score ability. The high score group gave the calibration of high score ability. The entire sample gave the whole range of ability calibration. The score-ability calibration curves obtained from the sample and various groups, are illustrated in the Figure 9. As it can be seen from the figure, they tend to increase with the same rate. The difference between the score-ability calibated from the low score group and that from the entire sample , was nearly a constant (approximately 0.9). The difference between the score-ability calibrated from the entire sample and that from the high score group, was equal to another constant(approximately 0.75). As it was predicted by Rasch(1966), the test exposed to different groups of respondents will give different solutions to the model parameters, which can differ by a constant value. Thus the finding seems to support Rasch's prediction.

TABLE 8

Ability of the score groups of the Test C as calibrated from the entire sample the high score group and the low score group

Score of score group	Low score group	High score group	total sample
(5)	-0.80511		-1.69580
(6)	-0.56279		-1.45442
(7)	-0.34776		-1.23979
(8)	-0.15218		-1.04412
(9)	0.02910		-0.86235
(10)	0.19972		-0.69089
(11)	0.36235		-0.52712
(12)	0.51908		-0.36897
(13)	0.67161		-0.21477
(14)		-0.81267	-0.06310
(15)		-0.65988	0.08735
(16)		-0.50689	0.23779
(17)		-0.35250	0.38945
(18)		-0.19542	0.54363
(19)		-0.03425	0.70176
(20)		0.13270	0.86552
(21)		0.30746	1.03696
(22)		0.49268	1.21873
(23)		0.69195	1.41441
(24)		0.91037	1.62909

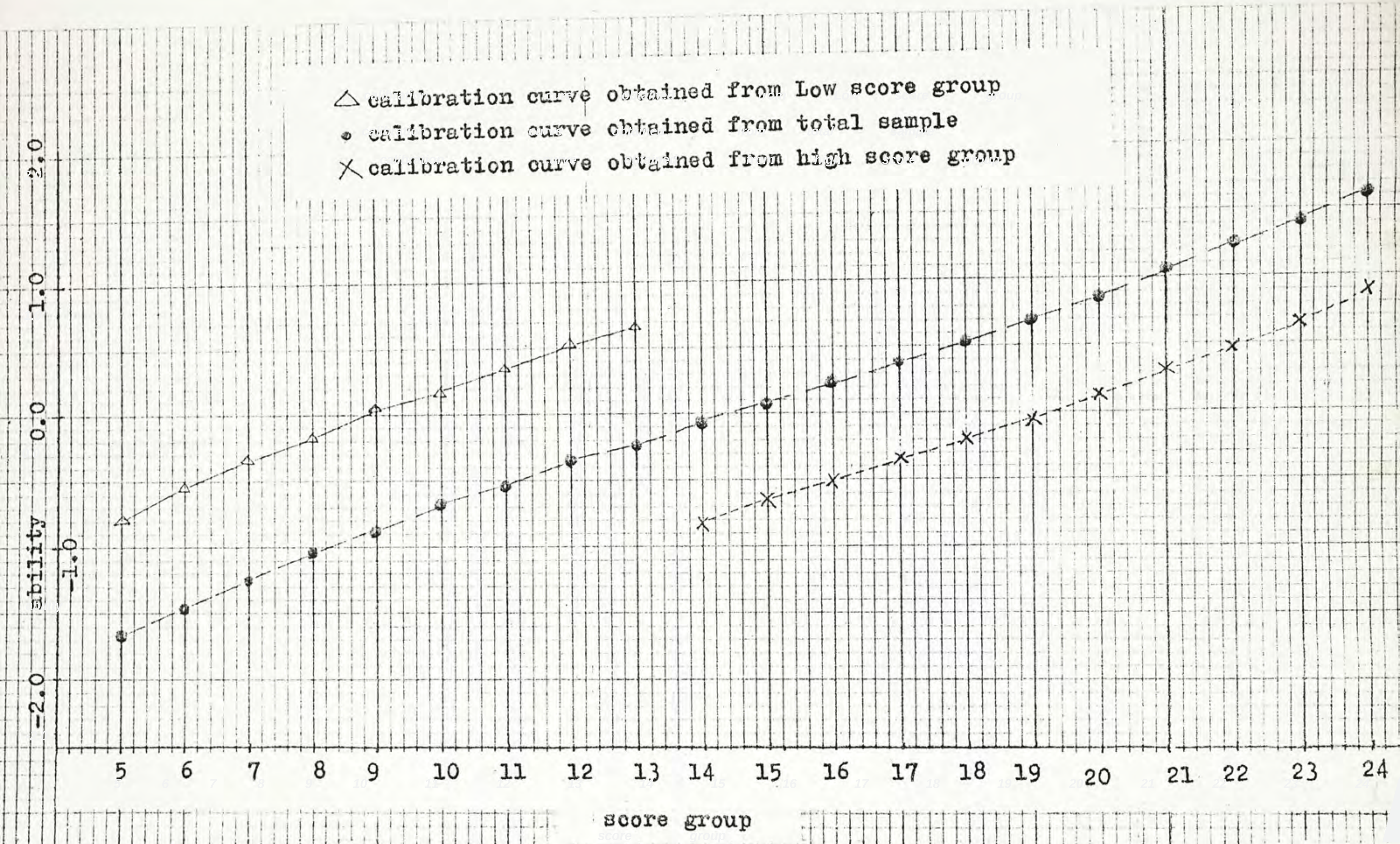


FIG. 9 Ability-score calibration curves obtained from the entire sample and different groups with the Test C

Therefore the hypothesis in this study, that there is no significant difference between item easiness as calibrated from the high score group and the low score group, is rejected. This finding can not rule out the objectivity of the Rasch model but it encourages us to define this property more clearly.

The Table 9 presents the correlation coefficient between the item easiness of the Rasch model , and the conventional item difficulty , as calibrated from different groups of respondents.

The item difficulty is defined as,

$$1 - \frac{n_i}{N}$$

where n_i = number of persons answering item i correctly

N = total number of persons in the sample or group considered.

Larger item easiness means smaller item difficulty, and greater probability to get the correct answer of the item i. As it can be seen from the Table 9, the results were approximately equal to the predicted value(r=1.0).

TABLE 9

correlation coefficients between item easiness
and item difficulty in the Test C as
calibrated from the entire sample
the high score group and
the low score group

	total sample	high score group	low score group
Correlation between item easiness and difficulty level	-0.9976	-0.9977	-0.9892

III. The invariance of person ability as calibrated from different test subsets

The ability for each subject was measured by the Hard test, and then the Easy test. The Hard test and the Easy test both were the subsets of the Test C.

The item characteristics and test of goodness of fit to the model are presented in the Table 10 and 11. There were only two items in the Hard test and one item in the Easy test, which did not fit the model in terms of chi square statistics.

The Figure 10 shows the point biserial correlation coefficients for the items in the Hard test and the Easy test. It seemed that they did not differ very much. They ranged from 0.3 to 0.5 .

The slope for each item in the two test subsets are demonstrated in the Figure 11. The major part of the items had slope in the range from 0.8 to 1.0. So that in general, we may conclude that most of the items also satisfy the criteria of fitting the model, introduced by Anderson(1968).

The Table 12 presents the reliability(Cronbach's coefficient alpha) and test of goodness of fit of the Test B, Test C, Hard test and the Easy test. The reliabilities of all tests were rather high. The Hard test and the Easy test seemed to have equal reliability, which was one of the requirements if they were considered as two equivalent tests. All tests, except the Test B, fitted the model in terms of chi square.

TABLE 10

Item characteristics and goodness of fit for each item in the Hard test as calibrated from the entire sample

Item No.	difficulty level	point biserial corr.	chi-square df= 9	slope df= 8
(1)	0.70474	0.34245	15.41436	1.05749
(2)	0.59267	0.49162	13.02811	0.76733 **
(3)	0.61207	0.42705	3.31876	0.86501
(4)	0.55819	0.37429	20.62656 *	1.08656
(5)	0.82112	0.40365	6.36809	0.91494
(6)	0.69397	0.38520	6.97327	0.93105
(7)	0.62284	0.43585	14.13791	0.80924
(8)	0.64224	0.40391	17.76320 *	0.90902
(9)	0.69181	0.34508	12.98054	1.03219
(10)	0.64009	0.41345	2.12810	0.95017
(11)	0.54741	0.52315	12.79597	0.67745 **
(12)	0.70905	0.41923	11.24590	1.03654
(13)	0.60991	0.53680	9.60507	0.68440 **
(14)	0.55388	0.39628	7.99974	1.07347
(15)	0.72414	0.44845	6.15176	0.66500 **

* significant in chi square at 0.05

** significant in departure from unit slope at 0.05

TABLE 11

Item characteristics and the goodness of fit for each item in the Easy test as calibrated from the entire sample

item No.	difficulty level	point biserial corr.	chi-square df= 11	slope df= 10
(1)	0.36588	0.29310	10.48589	1.29334 **
(2)	0.36284	0.33621	16.25846	1.26520
(3)	0.40131	0.42457	15.46800	1.15988
(4)	0.45945	0.37931	5.78542	0.94920
(5)	0.27479	0.21336	29.36354 *	1.02638
(6)	0.52703	0.45259	19.29629	0.82470
(7)	0.53646	0.38578	12.64961	0.78881 **
(8)	0.50614	0.44612	14.61874	0.78048 **
(9)	0.56589	0.42888	19.55986	0.61397 **
(10)	0.44133	0.32543	17.64552	0.94201
(11)	0.52853	0.47198	7.08406	0.84112 **
(12)	0.52519	0.43966	12.22237	0.80951
(13)	0.52031	0.41810	8.46057	0.78202 **
(14)	0.46988	0.51940	5.19697	0.84394
(15)	0.44697	0.53448	4.34172	1.05561

* significant in chi square at 0.05
** significant in departure from unit slope at 0.05

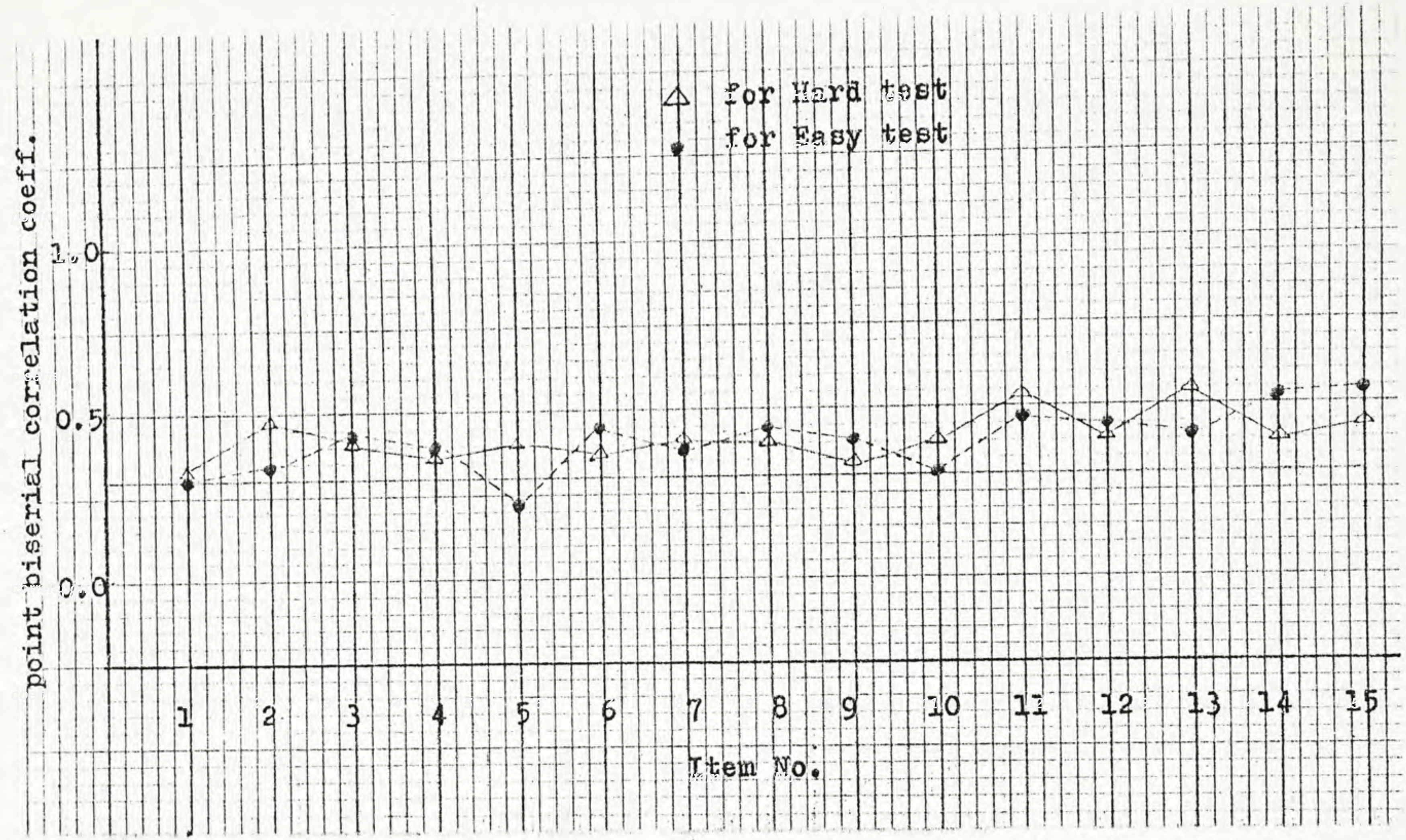


FIG. 10 Point biserial correlation coefficients for each item in the Hard test and the Easy test as calibrated from the entire sample

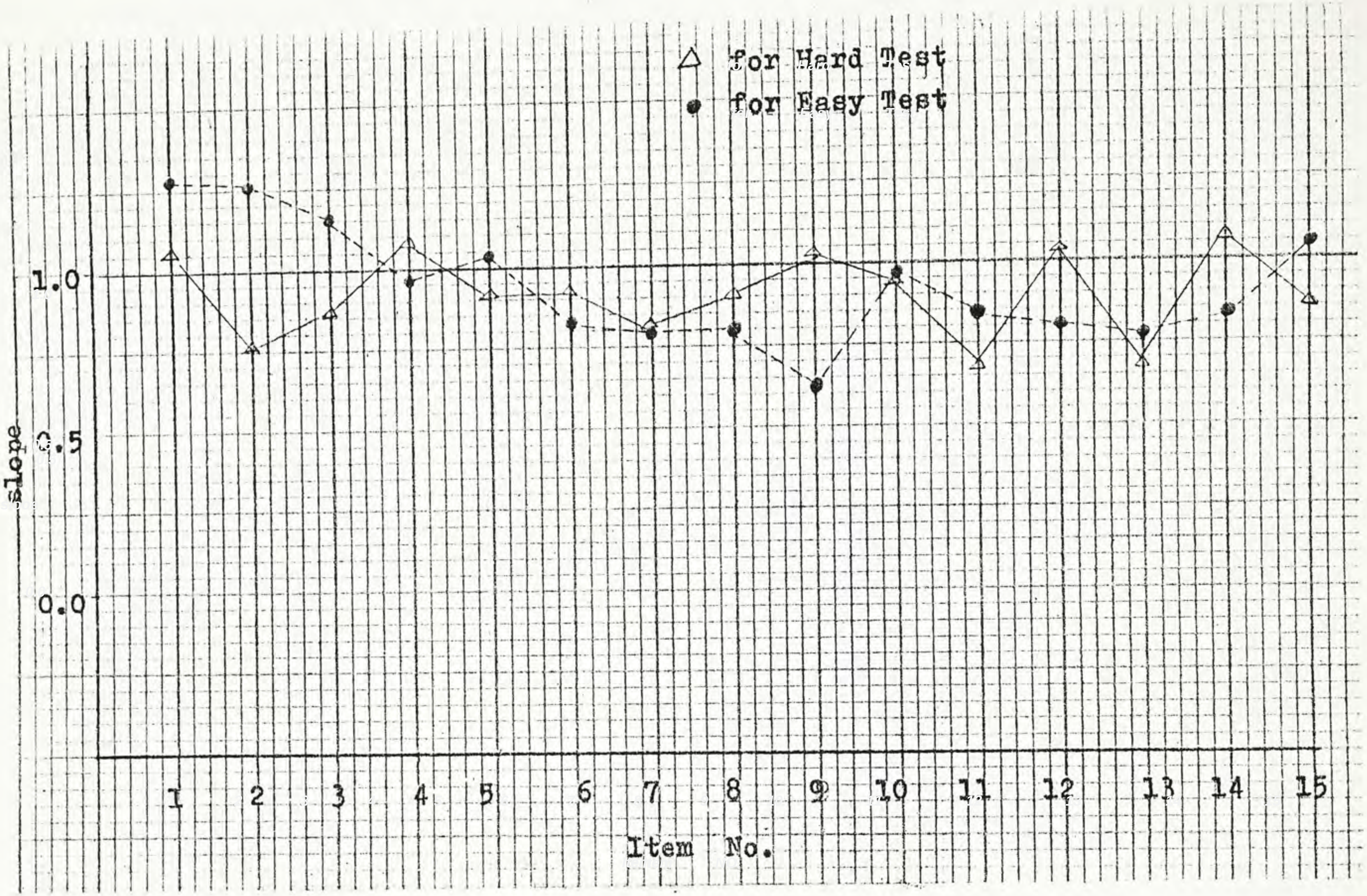


FIG. 11 Slope calculated for each item in the Hard test and the Easy test as calibrated in the entire sample

TABLE 12

The goodness of fit (in terms of chi square) and reliability(cronbach's coefficient α) of different test subsets

	Test B 47-item	Test C 30-item	Hard test 15-item	Easy Test 15-item
Reliability	0.819	0.820	0.673	0.741
Goodness of fit (χ^2)	2063.9**	629.9	160.5	198.4
df	855	551	126	154

** significant at 0.05

The Table 13 gives the conventional comparisions of ability of the sample as calibrated from two test subsets. The correlation coefficient between them was not very high, but significant at 0.05. The t test showed that there was no significant difference in the mean scores of person abilities.

The variance of the person abilities, as measured by two different test subsets, did not seem to differ very much on face value. But when they were tested by the Hartley's test, significant difference was found.

The Table 14 presents another type of data on the test of equivalence of the Hard test and the Easy test. The square of the standard error of measurement for two tests and the standardized difference scores were computed. The standard error of measurement seemed to be equal in the two tests. The mean of the standardized difference scores was very close to the expected value(0.0), but the standard deviation was very much larger than 1.0 .

Whitely(1974) reported the similar results, approximately zero in mean but significantly larger than one in standard deviation of the standardized difference scores.

Since this study did not produce two equivalent tests in the Rasch Model, the hypothesis that, there is no significant difference between ability calibrated from the Easy test and the Easy test, was also rejected.

TABLE 13

Comparis ons of person ability as calibrated from different test subsets

test subset	r	person ability $\bar{X}$	s	t	F_{max}
Easy test		-0.059	1.025		
v.s.	0.5174**			-1.123	1.44**
Hard test		-0.111	0.852		

** significant at 0.05

TABLE 14

Standardized difference scores of person ability as measured by different test subsets and the mean of standard error of measurement

Test subset	$^{a}\overline{SEM^2}$	standardized difference scores	
		$\bar{X}_1 - \bar{X}_2$	s
Easy test	0.01156	0.140	6.41
Hard test	0.01029		

a SEM^2 indicates the square of the standard error of measurement

CHAPTER FOUR

SUMMARY, CONCLUSION AND RECOMMENDATION

This study attempted to build up the objective property on a achievement test, through the Rasch-Wright's procedure. A test is objective if the calibration of the item parameters is independent of the sample and the calibration of person parameter is independent of the item subset.

A set of items fitting the Rasch model was selected to form the Test . The item parameters(easiness), of the the Test C, were calibrated from the high score group and low score group. The correlation coefficient between these two sets of item easiness was very high. Their mean difference was found to be significant. This is interesting for it seemed that the individual difference of the item easiness as calibrated from two groups of respondents, was a constant value. One set of item easiness could be made nearly identical to another set, by simply adding a constant value. This is permissible the Rasch model , as he pointed out that " the parameter η_s and ω_i are only determined up to an unspecified multiple" , in which η_s and ω_i are the ability and easiness (Rasch,1966,pp.50-51). When the parameters were changed to the logarithm scale in the Rasch-Wright procedure, they could only be determined up to an unspecified constant difference. As the constant difference is unspecified, its value is unknown until the two sets of item easiness have been calibrated from two different groups.

Because of the existence of this unspecified constant difference, a significant t test of the mean difference of two sets of item parameters is expected. In fact this is why Anderson(1968) did not employ the t test, but only used the correlation coefficient , in his study of independence of item parameters in the Rasch model.

Another interestin result was also obtained when the score group ability of the Test C as calibrated from the high score group, the low score group and the entire sample, were compared. Referring back to the Fighure 9, it seemed that constant differences existed among any two of the score-ability calibration curves. The three score-ability calibration curves can be adjusted to coincide with each other, by adding or subtracting a constant value . Thus it is possible to calibrate the score group ability of a test, from a sample of low ability, then from other sample of high ability separately. The combination of two score-ability calibration curves can give a calibration curve covering the whole range of scores. This is supported by Rasch, he said" the parameter of the subjects in the subgroup may be evaluated without regard to the parameters of other subjects"(Rasch, 1966,p.56).

Another phase of the objectivity of the Rasch model was also studied in this research. Ability for each subject in the sample was calibrated by the Hard test and the Easy test, which were subsets of the Test C. The correlation between the results of these two test subsets was moderately high. No significance in the t test of mean difference was found. However, when the standardized difference scores for these two sets of test were computed, they did not show that the Hard test and the Easy test were statistically equivalent. The standard deviation of these standardized difference scores deviated from 1.0 significantly. In other words, the ability difference scores obtained by the sample were not distributed as would be expected from the measurement error associated with each score(Whitely,1974).

This study failed to produce two equivalent tests. But this finding can not prove that the possibility is zero. Further studies on this problem should be encouraged. The following is a list of recommendations for further investigations.

1. As suggested by Rasch(1966), his model would fail to describe the data under the following conditions:

   a. when a single parameter for subject was not sufficient,
   b. when subjects were under the stress of time.

Usually, it is very difficult for an achievement test to confine the ability vector in one dimension. Cultural loading is another source introducing more dimensions in the ability vector.

2. During the construction of test items, the following points should be considered.

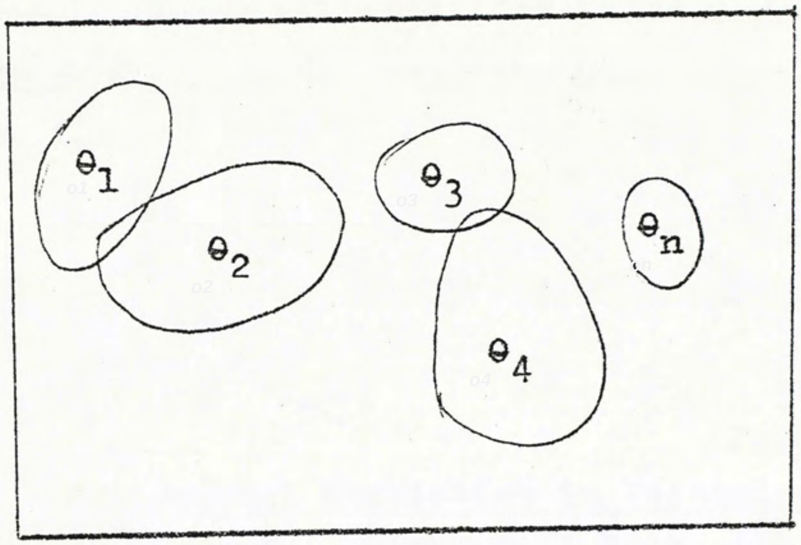
   a. maintaining high built-in validity,
   b. using words appropriate to subjects' language level,
   c. trying to use material not related with culture,
   d. keeping the difficulty index at medium level, in order to achieve minimum measurement and minimize the chance of introducing guessing factor.

3. When the test is administered to the sample, it is better to allow subjects with more time for response. Power test is a good way to relieve the time stressing.

4. The ability is not so simple that it can be separated into many dimensions. Usually in the real situation, the ability can be regarded as many regions. Some of them are separated, but some of them overlapp with each other. The idea is illustrated in the Figure 12. Because it is so difficult to separate the " ability region", it seems that construction of an unidimensional test is a problem in this kind of study.

5. Since there must be a constant difference between the item parameters calibrated from different samples, it is nearly impossible to obtain two sets of item parameters with no significant difference in mean and in variance. Thus the term" sample free" should be reviewed and possibly modified allowing the existence of the constant difference.

6. Whitely(1974) failed to produce two equivalent tests through the Rasch-Wrigh procedure. The criterion of testing the goodness of fit, in both this study and Whitely's study as the chi square statistics. This criterion can tell how well the model describes the data. Another criterion suggested by Anderson(1968), is the t test of unit slope for each item. This test can tell how well the model parameters can be separated. If the model parameters can be separated into person ability and item easiness, the slope calculated for each item will equal 1.0 . This test is more rigourous than the chi square statistics. Although the slope for each item has also been reported in this study, but it has not been used as the criterion for selecting fitted items. It is suggested that the criterin of unit slope should be used to repeat the study.



θ_i $i=1,2,\ldots\ldots n$
dimensions of the ability

FIG. 12 The imagined structure of ability

BIBLIOGRAPHY

Anderson, J., Kearney, G.E. and Everett, A.V., An evaluation of Rasch's structural model for test items, The British Journal of Mathematics and Statistical Psychology, 1968, 21, pp.231-238.

Draper, N.R., Smith, H. and Smith, N.N., Applied Regression Analysis, John Wiley, 1966.

Guilford, J.P., Fundamental Statistics in Psychology and Education, New York, Mcgraw-Hill Book, 1965.

Lord, F.M. and Novick, M.R. Statistical Theories of Mental Test Scores, Massachusetts, Adision-Wesley Publishing Company, 1968.

Rasch, G. An item analysis which takes individual difference into account, Statistical Psychology, 1966, 19, pp.49-57.

Whitely, S.E. and Dawis, R.V., The nature of objectivity with the Rasch model, Journal of Educational Measurement, 1975, 11, pp.163-178.

Wright, B. Sample free test calibration and person measurement, <u>Proceeding of the 1967 Invitation Conference on Testing Problems</u>, Princeton, N.J., Education Testing Service, 1967, pp. 85-101.

Wright, B. and Panchapakesan, N. A., A procedure for the sample free item analysis, <u>Educational and Psychological Measurement</u>, 1969, <u>29</u>, pp. 23-48.

Appendix A Programme Listings

(1) Programme of conventional analyses of items(reliability, point biserial corr. coeff., difficulty level, score mean and variance)

```
      MASTER MAK CHIU
C     FIRST PART OF THE PROGRAMME
      DIMENSION PB(50),TPB(50),NO(50),IX(50),DIFI(50),XX(50)
      COMMON X(500,30),Y(500)
      READ(1,10) N,K
10    FORMAT(2I3)
      WRITE(2,50) N,K
50    FORMAT(//2X,2I5)
C
C     SELECT ITEMS FOR ANALYSES
C
      DO 100 L=1,N
      READ(1,20)(XX(I),I=1,K)
20    FORMAT(50F1.0)
      X(L,1)=XX(4)
      X(L,2)=XX(5)
      X(L,3)=XX(6)
      X(L,4)=XX(7)
      X(L,5)=XX(9)
      X(L,6)=XX(10)
      X(L,7)=XX(11)
      X(L,8)=XX(12)
      X(L,9)=XX(13)
      X(L,10)=XX(18)
      X(L,11)=XX(19)
      X(L,12)=XX(20)
      X(L,13)=XX(21)
      X(L,14)=XX(22)
      X(L,15)=XX(24)
      X(L,16)=XX(25)
      X(L,17)=XX(26)
      X(L,18)=XX(28)
      X(L,19)=XX(29)
      X(L,20)=XX(30)
      X(L,21)=XX(31)
```

```
      X(L,22)=XX(35)
      X(L,23)=XX(36)
      X(L,24)=XX(37)
      X(L,25)=XX(39)
      X(L,26)=XX(40)
      X(L,27)=XX(41)
      X(L,28)=XX(43)
      X(L,29)=XX(45)
      X(L,30)=XX(48)
100   CONTINUE
C
      M=30
      DO 900 J=1,M
      NO(J)=J
900   CONTINUE
      WRITE(2,30)
30    FORMAT(1H1,///10X,15HRESPONSE MATRIX)
      DO 1111 I=1,M
      DO 2222 J=1,M
      IX(J)=X(I,J)
2222  CONTINUE
      WRITE(2,60)(IX(J),J=1,M)
60    FORMAT(10X,17I1)
1111  CONTINUE
C
      CALL STD(N,M,BARY,SY)
C
      WRITE(2,11) BARY,SY
11    FORMAT(///10X,15HMEAN OF SCORE =,F8.3,///10X,
     214HSTD OF SCORE =,F8.3)
C
      DO 500 J=1,M
      CALL RPB(J,BARY,SY,N,PB,TPB)
500   CONTINUE
C
C
```

```
C     CALCULATION OF DIFFICULTY LEVEL FOR EACH ITEM
C
      XN=N
      DO 200 J=1,M
      DIFI(J)=0.0
      DO 210 I=1,N
      DIFI(J)=DIFI(J)+X(I,J)
210   CONTINUE
      DIFI(J)=1.0-(DIFI(J)/XN)
200   CONTINUE
C
      WRITE(2,22)
22    FORMAT(////10X,97HPOINT BISERIAL CORRELATION FOR ITEMS              T
     3ST OF SIGNIFICANCE     DIFFICULTY LEVEL FOR ITEMS)
      WRITE(2,24)(NO(J),PB(J),TPB(J),DIFI(J),J=1,M)
24    FORMAT(5X,1H(,I3,1H),F10.5,35X,F10.5,17X,F10.5)
C
      CALL CRONBK(N,M,ALPHA)
C
      WRITE(2,31) ALPHA
31    FORMAT(///2X,39HRELIABILITY(CRONBACK ALPHA) OF TEST IS=,F10.6)
C
      CALL XKUDERR(SY,N,M,XKRR)
C
      WRITE(2,41) XKRR
41    FORMAT(///2X,31HRELIABILITY(KR=20) OF TEST IS =,F10.6)
C
      CALL RANKORDER(N)
C
      WRITE(2,12)
12    FORMAT(///2X,14HSCORE IN ORDER)
      WRITE(2,13)(Y(I),I=1,N)
13    FORMAT(2X,10F10.1)
      STOP
      END
```

```
      SUBROUTINE STD(N,M,BARY,SY)
      COMMON X(500,30),Y(500)
      DO 100 I=1,N
      Y(I)=0.0
100   CONTINUE
      DO 200 I=1,N
      DO 200 J=1,M
      Y(I)=X(I,J)+Y(I)
200   CONTINUE
      DO 400 I=1,N
      TA=Y(I)**2+TA
      TB=Y(I)+TB
400   CONTINUE
      SN=N
      BARY=TB/SN
      SY=SQRT((TA-(TB**2)/SN)/(SN-1.0))
      RETURN
      END
```

```
      SUBROUTINE RPB(J,BARY,SY,N,PB,TPB)
      DIMENSION PB(50),TPB(50)
      COMMON X(500,30),Y(500)
      YA=0.0
      YB=0.0
      SNA=0.0
      SNB=0.0
      DO 200 I=1,N
      IF(X(I,J)-1.0) 1,2,2
1     SNA=SNA+1.0
      YA=YA+Y(I)
      GO TO 200
2     SNB=SNB+1.0
      YB=YB+Y(I)
200   CONTINUE
      SN=N
      IF(SNA.EQ.SN) GO TO 3
      IF(SNB.EQ.SN) GO TO 4
      BARYA=YA/SNA
      BARYB=YB/SNB
      GO TO 5
3     BARYA=YA/SNA
      BARYB=0.0
      GO TO 5
4     BARYB=YB/SNB
      BARYA=0.0
5     TB=SQRT((SNB*SNA)/(SN*(SN-1.0)))
      PB(J)=((BARYB-BARYA)/SY)*TB
      TPB(J)=(PB(J)*SQRT(SN-2.0))/(SQRT(1.0-(PB(J)**2)))
      RETURN
      END
```

```
      SUBROUTINE CRONBK(N,M,ALPHA)
      DIMENSION TB(500),TC(50)
      COMMON X(500,30),Y(500)
      A=0.0
      B=0.0
      C=0.0
      D=0.0
      DO 100 J=1,M
100   TC(J)=0.0
      DO 200 I=1,N
200   TB(I)=0.0
      DO 300 I=1,N
      DO 310 J=1,M
      D=D+X(I,J)
      TB(I)=TB(I)+X(I,J)
      A=A+(X(I,J))**2
310   CONTINUE
      TB(I)=(TB(I))**2
      B=B+TB(I)
300   CONTINUE
      D=D**2
      DO 400 J=1,M
      DO 410 I=1,N
      TC(J)=TC(J)+X(I,J)
410   CONTINUE
      TC(J)=(TC(J))**2
      C=C+TC(J)
400   CONTINUE
      XM=M
      XN=N
      B=B/XM
      C=C/XN
      D=D/(XN*XM)
      G=(A-B-C+D)/((XN-1.0)*(XM-1.0))
      H=(B-D)/(XN-1.0)
      ALPHA=1.0-G/H
      RETURN
      END
```

```
      SUBROUTINE XKUDERR(SY,N,M,XKRR)
      DIMENSION XP(50)
      COMMON X(500,30),Y(500)
      XN=N
      DO 100 J=1,M
      XP(J)=0.0
      DO 110 I=1,N
      XP(J)=XP(J)+X(I,J)
110   CONTINUE
      XP(J)=XP(J)
      XP(J)=XP(J)/XN
100   CONTINUE
      A=0.0
      DO 200 J=1,M
      A=A+XP(J)*(1.0-XP(J))
200   CONTINUE
      XM=M
      XKRR=(XM/(XM-1.0))*((SY**2-A)/SY**2)
      RETURN
      END
```

```
      SUBROUTINE RANKORDER(N)
      COMMON X(500,30),Y(500)
      M=N-1
      DO 100 I=1,M
      L=I+1
      DO 110 J=L,N
      IF(Y(I)-Y(J)) 1,110,110
1     TEMP=Y(I)
      Y(I)=Y(J)
      Y(J)=TEMP
110   CONTINUE
100   CONTINUE
      RETURN
      END
```

```
PROGRAM JUCC
EXTENDED DATA (22AM)
COMPACT PROGRAM (DBM)
CORE              38592

              SEG     MAKCHIU
              SEG     STD
              SEG     RPB
              SEG     CRONBK
              SEG     XKUDERR
              SEG     RANKORDER
              SEG     SQRT
```

(2) Programme of Rasch-Wright procedure of item analyses

```
      MASTER MAK CHIU
C     SECOND PART OF THE PROGRAMME
      DIMENSION X(50),AP(50,50),RP(50),SCR(50),NO(50),B(50),XCHISQ(50),
     1IX(50),TEST(50),TSCR(500),CR(50),SRR(50),IAP(50),SEIT(50),SFAB(50)
      COMMON Y(500),AA(30,30),RR(30),T(30,30),EAS(30),ABL(30)
      READ(1,10) N,K
10    FORMAT(2I3)
      WRITE(2,50) N,K
50    FORMAT(//2X,2I5)
      M=30
      DO 100 I=1,(M-1)
      RP(I)=0.0
      DO 110 J=1,M
      AP(I,J)=0.0
110   CONTINUE
100   CONTINUE
      DO 900 J=1,M
      NO(J)=J
900   CONTINUE
C
C    SELECT ITEMS WHICH FIT THE MODEL FOR ANALYSES
C     TRANSFORM SUBJECTS REPONSES TO MATRIX A AND R
C
      DO 120 I=1,N
      READ(1,20)(X(J),J=1,K)
20    FORMAT(50F1.0)
      X(1)=X(4)
      X(2)=X(5)
      X(3)=X(6)
      X(4)=X(7)
      X(5)=X(9)
      X(6)=X(10)
      X(7)=X(11)
      X(8)=X(12)
      X(9)=X(13)
      X(10)=X(18)
      X(11)=X(19)
```

```
      X(12)=X(20)
      X(13)=X(21)
      X(14)=X(22)
      X(15)=X(24)
      X(16)=X(25)
      X(17)=X(26)
      X(18)=X(28)
      X(19)=X(29)
      X(20)=X(30)
      X(21)=X(31)
      X(22)=X(35)
      X(23)=X(36)
      X(24)=X(37)
      X(25)=X(39)
      X(26)=X(40)
      X(27)=X(41)
      X(28)=X(43)
      X(29)=X(45)
      X(30)=X(48)
      DO 2222 JJ=1,M
      IX(JJ)=X(JJ)
2222  CONTINUE
      WRITE(2,23)(IX(JJ),JJ=1,M)
23    FORMAT(10X,30I1)
      Y(I)=0.0
      TEMP=0.0
      DO 210 LL=1,M
      TEMP=TEMP+X(LL)
210   CONTINUE
      Y(I)=TEMP
      L=Y(I)
      IF(L) 120,120,1
1     IF(L.EQ.M) GO TO 120
      RP(L)=RP(L)+1.0
      TS=TS+1.0
```

```
      DO 220 J=1,M
      AP(L,J)=AP(L,J)+X(J)
220   CONTINUE
120   CONTINUE
      WRITE(2,51) TS
51    FORMAT(///10X,32HTOTAL STUDENT BEING ANALIZED IS=,F10.6)
      WRITE(2,52)
52    FORMAT(///10X,8HMATRIX R)
      WRITE(2,53)(RP(I),I=1,(M-1))
53    FORMAT(10X,F10.2)
      WRITE(2,54)
54    FORMAT(///10X,8HMATRIX A)
      DO 3333 I=1,(M-1)
      DO 4444 J=1,M
      IAP(J)=AP(I,J)
4444  CONTINUE
      WRITE(2,15)(IAP(J),J=1,M)
15    FORMAT(2X,30I3)
3333  CONTINUE
C
C
      DO 1110 I=1,(M-1)
      CR(I)=I
1110  CONTINUE
C     DELETE THOSE SCORE GROUP WITH RESPONDENT LESS THAN 10
      DO 1111 I=1,(M-1)
      IF(RP(I).LT.10.0) GO TO 121
      GO TO 1111
121   RP(I)=0.0
1111  CONTINUE
C
```

```
C      CLOSE VECTOR R AND MATRIX A
C
       NK=0
       DO 300 I=1,(M-1)
       IF(RP(I).EQ.0.0) GO TO 300
       NK=NK+1
300    CONTINUE
       WRITE(2,38) NK
38     FORMAT(///2X,29HTOTAL NUMBER OF ABILITY GROUP,/2X,I3)
       DO 320 I=1,NK
       DO 330 J=1,M
       AA(I,J)=0.0
330    CONTINUE
320    CONTINUE
       DO 400 J=1,NK
       RR(J)=0.0
       DO 410 I=1,(M-1)
       IF(RP(I).EQ.0.0) GO TO 410
       IF(RR(J).EQ.0.0) GO TO 4
       GO TO 410
4      RR(J)=RP(I)
       RP(I)=0.0
       SRR(J)=CR(I)
       DO 420 K=1,M
       AA(J,K)=AP(I,K)
420    CONTINUE
410    CONTINUE
400    CONTINUE
       WRITE(2,32)
32     FORMAT(//2X,24HSCORE OF THE SCORE GROUP)
       WRITE(2,31) (SRR(I),I=1,NK)
31     FORMAT(2X,F10.1)
       WRITE(2,34)
34     FORMAT(///2X,17HMATRIX R (CLOSED))
       WRITE(2,35)(RR(I),I=1,NK)
35     FORMAT(2X,F10.2)
```

```
      WRITE(2,36)
36    FORMAT(///2X,17HMATRIX A (CLOSED))
      DO 3334 I=1,NK
      DO 4445 J=1,M
      IAP(J)=AA(I,J)
4445  CONTINUE
      WRITE(2,17)(IAP(J),J=1,M)
17    FORMAT(2X,30I3)
3334  CONTINUE
C
C
C     ASSIGN VALUE TO THE TOTAL SCORE OF EACH PERSON
C
      DO 710 I=1,N
      TSCR(I)=Y(I)
710   CONTINUE
C
C
      CALL TMATRIX(NK,M)
C
      WRITE(2,61)
61    FORMAT(///10X,8HMATRIX T)
      WRITE(2,62)((T(I,J),J=1,M),I=1,NK)
62    FORMAT(///2X,10F10.5,/2X,10F10.5,/2X,10F10.5)
C
      CALL XLOGEST(NK,M)
C
      WRITE(2,71)
71    FORMAT(///10X,22HABILITY OF SCORE GROUP)
      WRITE(2,72)(NO(I),ABL(I),I=1,NK)
72    FORMAT(5X,1H(,I3,1H),2X,F10.5)
      WRITE(2,73)
73    FORMAT(///10X,17HEASINESS OF ITEMS)
      WRITE(2,74)(NO(J),EAS(J),J=1,M)
74    FORMAT(5X,1H(,I3,1H),2X,F10.5)
C
```

```
      CALL EMFUCELL(NK,M)
C
      WRITE(2,81)
81    FORMAT(///10X,17HMODIFIED MATRIX T)
      WRITE(2,82)((T(I,J),J=1,M),I=1,NK)
82    FORMAT(///2X,10F10.5,/2X,10F10.5,/2X,10F10.5)
C
      CALL XLOGEST(NK,M)
C
      WRITE(2,75)
75    FORMAT(///10X,22HABILITY OF SCORE GROUP)
      WRITE(2,76)(NO(I),ABL(I),I=1,NK)
76    FORMAT(5X,1H(,I3,1H),2X,F10.5)
      WRITE(2,77)
77    FORMAT(///10X,17HEASINESS OF ITEMS)
      WRITE(2,78)(NO(J),EAS(J),J=1,M)
78    FORMAT(5X,1H(,I3,1H),2X,F10.5)
      WRITE(2,92)
92    FORMAT(///10X,20HPROCESS OF ITERATION)
C
      CALL XMAXITERAT(NK,M,SRR)
C
      WRITE(2,93)
93    FORMAT(///10X,19HRESULT OF ITERATION)
      WRITE(2,96)
96    FORMAT(//10X,22HABILITY OF SCORE GROUP)
      WRITE(2,94)(NO(I),ABL(I),I=1,NK)
94    FORMAT(5X,1H(,I3,1H),F10.5)
      WRITE(2,97)
97    FORMAT(//10X,17HEASINESS OF ITEMS)
      WRITE(2,95)(NO(J),EAS(J),J=1,M)
95    FORMAT(5X,1H(,I3,1H),F10.5)
```

```
C
      CALL FITNESS(NK,M,XCHISQ,CHISQ,SEIT,SEAB)
C
      WRITE(2,90)
90    FORMAT(////2X,29HCHISQUARE FOR INDIVIDUAL ITEM)
      WRITE(2,70)(NO(J),XCHISQ(J),J=1,M)
70    FORMAT(5X,1H(,I3,1H),2X,F10.5)
      WRITE(2,80) CHISQ
80    FORMAT(///2X,25HCHISQUARE FOR WHOLE TEST=,F10.5)
      WRITE(2,84)
84    FORMAT(/////5X,44HABILITY         STANDARD ERROR OF MEASUREMENT)
      WRITE(2,85)(ABL(I),SEAB(I),I=1,NK)
85    FORMAT(5X,F10.5,5X,F10.5)
      WRITE(2,86)
86    FORMAT(/////5X,44HEASINESS        STANDARD ERROR OF MEASUREMENT)
      WRITE(2,87)(EAS(J),SEIT(J),J=1,M)
87    FORMAT(5X,F10.5,5X,F10.5)
C
      CALL CRITB(NK,M,B,TEST)
C
      WRITE(2,91)
91    FORMAT(///10X,47HSLOPE FOR EACH ITEM              TEST OF UNIT SLOPE)
      WRITE(2,99)(NO(J),B(J),TEST(J),J=1,M)
99    FORMAT(5X,1H(,I3,1H),2X,F10.5,8X,F10.5)
C
C
C     CALIBRATION OF PERSON ABILITY
C
C
      WRITE(2,101)
101   FORMAT(///2X,29HCALIBRATION OF PERSON ABILITY)
      WRITE(2,102)
102   FORMAT(//2X,50HSUBJECT     TOTAL SCORE     SCORE GROUP    ABILITY)
      XM=M
      DO 700 I=1,N
      NNO=I
      XXX=0.0
      IF(TSCR(I).EQ.0.0) GO TO 8
      IF(TSCR(I).EQ.XM) GO TO 8
      DO 800 K=1,NK
      IF(TSCR(I).EQ.SRR(K)) GO TO 7
      GO TO 800
7     WRITE(2,104) NNO,TSCR(I),SRR(K),ABL(K)
104   FORMAT(2X,1H(,I3,1H),8X,F10.4,7X,F10.4,4X,F10.4)
```

```
      XXX=1.0
800   CONTINUE
      IF(XXX.EQ.1.0) GO TO 700
      WRITE(2,105) NNO,TSCR(I)
105   FORMAT(2X,1H(,I3,1H),8X,F10.4,7X,33HSCORE NOT IN THE CALIBRATED GR
     2OUP)
      GO TO 700
8     WRITE(2,103) NNO,TSCR(I)
103   FORMAT(2X,1H(,I3,1H),8X,F10.4,7X,28HABILITY CANNOT BE DETERMINED)
700   CONTINUE
      STOP
      END


      SUBROUTINE TMATRIX(N,M)
      COMMON Y(500),A(50,50),R(50),T(50,50),EAS(50),ABL(50)
      TS=0.0
      DO 100 I=1,N
100   TS=TS+R(I)
      DO 200 I=1,N
      DO 300 J=1,M
      IF(R(I).EQ.0.0) GO TO 1
      IF(A(I,J)) 2,2,3
3     IF(A(I,J).EQ.R(I)) GO TO 2
      T(I,J)=ALOG(A(I,J)/(R(I)-A(I,J)))
      GO TO 300
2     T(I,J)=ALOG((A(I,J)+R(I)/TS)/(R(I)-A(I,J)+R(I)/TS))
      GO TO 300
1     T(I,J)=0.0
300   CONTINUE
200   CONTINUE
      RETURN
      END
```

```
      SUBROUTINE XMAXITERAT(N,M,SCR)
      DIMENSION SCR(50),YG(50),GPY(50),XF(50),YF(50),FPY(50),F(50),G(50)
      COMMON Y(500),A(30,30),R(30),T(30,30),EAS(30),ABL(30)
      DO 900 K=1,20
      DO 200 I=1,N
      YG(I)=0.0
      GPY(I)=0.0
200   CONTINUE
      DO 300 J=1,M
      XF(J)=0.0
      YF(J)=0.0
      FPY(J)=0.0
300   CONTINUE
      DO 400 J=1,M
      DO 410 I=1,N
      XF(J)=XF(J)+A(I,J)
      YF(J)=YF(J)+R(I)*(EXP(EAS(J)+ABL(I)))/(1+EXP(EAS(J)+ABL(I)))
      FPY(J)=FPY(J)+R(I)*(EXP(EAS(J)+ABL(I)))/(1+EXP(EAS(J)+ABL(I)))**2
410   CONTINUE
      F(J)=XF(J)-YF(J)
      EAS(J)=EAS(J)+F(J)/FPY(J)
400   CONTINUE
      DO 500 I=1,N
      DO 510 J=1,M
      YG(I)=YG(I)+(EXP(ABL(I)+EAS(J)))/(1+EXP(ABL(I)+EAS(J)))
      GPY(I)=GPY(I)+(EXP(ABL(I)+EAS(J)))/(1+EXP(ABL(I)+EAS(J)))**2
510   CONTINUE
      G(I)=SCR(I)-YG(I)
      ABL(I)=ABL(I)+G(I)/GPY(I)
500   CONTINUE
      WRITE(2,70)(EAS(J),J=1,M)
70    FORMAT(///2X,10F10.5,/2X,10F10.5,/2X,10F10.5)
      WRITE(2,80)(ABL(I),I=1,N)
80    FORMAT(/2X,10F10.5)
900   CONTINUE
      RETURN
      END
```

```
      SUBROUTINE XLOGEST(N,M)
      DIMENSION TMR(50),TMC(50)
      COMMON Y(500),A(50,50),R(50),T(50,50),EAS(50),ABL(50)
      XM=M
      XN=N
      TTM=0.0
      DO 100 I=1,N
      TMR(I)=0.0
      DO 110 J=1,M
  110 TTM=TTM+T(I,J)
  100 CONTINUE
      TTM=TTM/(XN*XM)
      DO 200 J=1,M
  200 TMC(J)=0.0
      DO 300 I=1,N
      DO 310 J=1,M
  310 TMR(I)=TMR(I)+T(I,J)
      TMR(I)=TMR(I)/XM
  300 ABL(I)=TMR(I)-TTM
      DO 400 J=1,M
      DO 410 I=1,N
  410 TMC(J)=TMC(J)+T(I,J)
      TMC(J)=TMC(J)/XN
  400 EAS(J)=TMC(J)-TTM
      RETURN
      END
```

```
      SUBROUTINE FITNESS(N,M,XCHISQ,CHISQ,SEIT,SEAB)
      DIMENSION XCHISQ(50),P(50,50),VAR(50,50),Z(50,50),SEIT(50),
     1SEAB(50)
      COMMON Y(500),A(30,30),R(30),T(30,30),EAS(30),ABL(30)
      DO 100 I=1,N
      DO 110 J=1,M
      P(I,J)=(EXP(ABL(I)+EAS(J)))/(1.0+EXP(ABL(I)+EAS(J)))
      VAR(I,J)=R(I)*P(I,J)*(1.0-P(I,J))
      Z(I,J)=(A(I,J)-R(I)*P(I,J))/(SQRT(VAR(I,J)))
110   CONTINUE
100   CONTINUE
      DO 210 J=1,M
      XCHISQ(J)=0.0
210   CONTINUE
      CHISQ=0.0
      DO 300 J=1,M
      DO 310 I=1,N
      XCHISQ(J)=XCHISQ(J)+(Z(I,J))**2
310   CONTINUE
      CHISQ=CHISQ+XCHISQ(J)
300   CONTINUE
      XN=N
      XM=M
      DO 400 J=1,M
      W=0.0
      DO 410 I=1,N
      W=W+1/(R(I)*P(I,J)*(1.0-P(I,J)))
410   CONTINUE
      SEIT(J)=W/(XN**2)
400   CONTINUE
      DO 500 I=1,N
      W=0.0
      DO 510 J=1,M
      W=W+1/(R(I)*P(I,J)*(1.0-P(I,J)))
510   CONTINUE
      SEAB(I)=W/(XM**2)
500   CONTINUE
      RETURN
      END
```

```
      SUBROUTINE CRITR(N,M,B,TEST)
      DIMENSION B(50),TMR(50),TEST(50)
      COMMON Y(500),A(50,50),R(50),T(50,50),EAS(50),ABL(50)
      XM=M
      XN=N
      DO 100 I=1,N
      TMR(I)=0.0
      DO 110 J=1,M
      TMR(I)=TMR(I)+T(I,J)
110   CONTINUE
      TMR(I)=TMR(I)/XM
100   CONTINUE
      DO 200 J=1,M
      F=0.0
      E=0.0
      C=0.0
      D=0.0
      G=0.0
      DO 210 I=1,N
      F=F+T(I,J)*TMR(I)
      E=E+T(I,J)
      C=C+TMR(I)
      D=D+(T(I,J))**2
      G=G+(TMR(I))**2
210   CONTINUE
      B(J)=(F-(E*C)/XN)/(D-(E**2)/XN)
      DM=SQRT(D-(E**2)/XN)
      AA=G-(C**2)/XN
      BB=B(J)*(F-(E*C)/XN)
      SS=AA-BB
      S=SQRT(SS/(XN-2.0))
      TEST(J)=((B(J)-1.0)*DM)/S
200   CONTINUE
      RETURN
      END
```

```
      SUBROUTINE EMFUCELL(N,M)
      COMMON Y(500),A(50,50),R(50),T(50,50),EAS(50),ABL(50)
      DO 100 I=1,N
      DO 110 J=1,M
      IF(A(I,J).EQ.0.0) GO TO 999
      IF(A(I,J).EQ.R(I)) GO TO 999
      T(I,J)=T(I,J)
      GO TO 110
999   T(I,J)=ABL(I)+EAS(J)
110   CONTINUE
100   CONTINUE
      RETURN
      END
```

```
PROGRAM JUCC
EXTENDED DATA (22AM)
COMPACT PROGRAM (DBM)
CORE              42304

            SEG       MAXCHIU
            SEG       TMATRIX
            SEG       XLOGEST
            SEG       EMFUCELL
            SEG       XMAXITERAT
            SEG       FITNESS
            SEG       CRIT8
            SEG       ALOG
            SEG       EXP
            SEG       SQRT
```

(3) Programme output of the Rasch-Wright procedure

```
TOTAL STUDENT BEING ANALIZED IS=461.000000


MATRIX R
       0.00
       3.00
       4.00
       9.00
      11.00
      15.00
      17.00
      20.00
      20.00
      25.00
      30.00
      37.00
      32.00
      29.00
      28.00
      27.00
      23.00
      19.00
      21.00
      16.00
      20.00
      17.00
      14.00
      10.00
       5.00
       2.00
       6.00
       0.00
       1.00
```

MATRIX A

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	2	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	2	0	0	0	0	0	0	1	2	0	1	2	0	0	0	0	0	0	1	0	0	0
3	3	0	2	0	3	0	3	6	0	0	0	1	0	0	3	1	2	0	5	0	2	1	0	0	1	0	0	0	0
2	5	1	2	0	4	3	4	6	0	0	1	1	1	1	2	1	0	1	5	3	2	1	1	3	0	1	2	2	0
1	7	3	8	2	4	3	6	11	0	2	3	1	2	3	3	3	3	1	6	2	4	3	1	2	1	2	1	2	0
2	9	3	9	5	6	2	3	9	1	7	2	5	2	7	4	6	5	3	3	3	5	3	1	1	0	5	3	5	0
3	9	6	7	5	8	6	9	13	1	2	2	4	7	7	7	5	7	5	11	2	6	2	3	8	4	4	4	1	2
4	8	3	8	3	9	9	10	17	2	4	4	3	4	8	12	6	4	4	12	4	5	2	3	7	4	6	8	3	4
3	17	8	10	9	12	10	12	18	4	6	11	4	3	7	9	8	10	8	13	11	4	7	7	10	5	7	6	10	1
6	17	10	19	8	13	9	16	23	3	2	6	5	10	10	14	13	10	7	19	15	16	10	9	14	10	10	10	11	5
10	23	12	26	14	19	12	21	30	4	9	14	15	5	18	19	19	17	9	20	17	14	12	5	21	4	18	15	15	7
4	23	5	20	8	18	11	19	26	7	9	14	8	11	15	22	18	20	11	23	15	20	13	6	14	9	14	15	10	8
14	19	14	23	12	18	13	17	21	3	9	8	8	4	15	17	15	20	5	20	18	15	15	0	17	13	6	17	16	14
7	25	11	17	11	17	18	16	23	5	10	14	11	9	18	15	16	15	10	16	19	17	11	9	19	12	15	11	16	7
8	21	11	22	14	18	14	19	23	5	9	13	10	7	19	19	19	21	10	22	18	16	16	7	14	10	15	12	12	8
5	20	9	18	11	11	13	14	21	3	7	12	9	12	16	19	18	18	14	19	15	20	12	7	17	10	10	12	14	5
8	17	9	14	8	14	10	15	15	3	11	4	7	8	16	14	13	14	12	15	15	15	9	7	16	10	11	16	9	7
7	17	15	19	9	16	14	19	16	6	2	4	14	10	18	18	15	15	8	16	13	17	15	12	17	16	15	15	14	7
9	15	8	14	10	10	10	13	15	1	4	11	6	7	13	14	15	15	9	14	9	13	12	10	15	10	11	10	12	5
9	16	15	16	14	17	11	18	18	6	11	7	14	7	18	20	16	17	10	19	17	14	17	8	19	12	13	16	12	13
10	15	14	15	7	15	10	15	17	6	6	11	10	8	14	16	13	17	10	16	14	17	11	11	17	12	13	12	15	7
8	14	11	12	10	12	8	14	12	5	11	9	7	11	9	11	12	12	9	13	10	13	13	8	14	13	7	12	13	9
4	9	7	9	7	7	6	10	8	4	8	9	8	4	7	10	8	7	6	9	9	10	9	8	10	10	9	10	10	8
0	5	4	5	3	5	3	4	4	5	5	5	5	2	5	5	4	4	4	5	5	5	5	3	5	4	4	5	4	3
1	2	1	2	2	2	2	2	2	0	2	2	1	1	2	2	2	2	2	1	2	2	2	2	2	2	1	2	2	2
4	6	6	6	5	6	5	6	6	6	3	6	6	5	5	6	6	6	5	6	6	5	6	4	5	6	6	6	5	3
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

TOTAL NUMBER OF ABILITY GROUP
20

SCORE OF THE SCORE GROUP
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
13.0
14.0
15.0
16.0
17.0
18.0
19.0
20.0
21.0
22.0
23.0
24.0

MATRIX R (CLOSED)

11.00
15.00
17.00
20.00
20.00
25.00
30.00
37.00
32.00
29.00
28.00
27.00
23.00
19.00
21.00
16.00
20.00
17.00
14.00
10.00

MATRIX A (CLOSED)

2	5	1	2	0	4	3	4	6	0	0	1	1	1	1	2	1	0	1	5	3	2	1	1	3	0	1	2	2	0
1	7	3	8	2	4	3	6	11	0	2	3	1	2	3	3	3	3	1	6	2	4	3	1	2	1	2	1	2	0
2	9	3	9	5	6	2	3	9	1	7	2	5	2	7	4	6	5	3	3	3	5	3	1	1	0	5	3	5	0
3	9	6	7	5	8	6	9	13	1	2	2	4	7	7	7	5	7	5	11	2	6	2	3	8	4	4	4	1	2
4	8	3	8	3	9	9	10	17	2	4	4	3	4	8	12	6	4	4	12	4	5	2	3	7	4	6	8	3	4
3	17	8	10	9	12	10	12	18	4	6	11	4	3	7	9	8	10	8	13	11	4	7	7	10	5	7	6	10	1
6	17	10	19	8	13	9	16	23	3	2	6	5	10	10	14	13	10	7	19	15	16	10	9	14	10	10	10	11	5
10	23	12	26	14	19	12	21	30	4	9	14	15	5	18	19	19	17	9	20	17	14	12	5	21	4	18	15	15	7
4	23	5	20	8	18	11	19	26	7	9	14	8	11	15	22	18	20	11	23	15	20	13	6	14	9	14	15	10	8
14	19	14	23	12	18	13	17	21	3	9	8	8	4	15	17	15	20	5	20	18	15	15	0	17	13	6	17	16	14
7	25	11	17	11	17	18	16	23	5	10	14	11	9	18	15	16	15	10	16	19	17	11	9	19	12	15	11	16	7
8	21	11	22	14	18	14	19	23	5	9	13	10	7	19	19	19	21	10	22	18	16	16	7	14	10	15	12	12	8
5	20	9	18	11	11	13	14	21	3	7	12	9	12	16	19	18	18	14	19	15	20	12	7	17	10	10	12	14	5
8	17	9	14	8	14	10	15	15	3	11	4	7	8	16	14	13	14	12	15	15	15	9	7	16	10	11	16	9	7
7	17	15	19	9	16	14	19	16	6	2	4	14	10	18	18	15	15	8	16	13	17	15	12	17	16	15	15	14	7
9	15	8	14	10	10	10	13	15	1	4	11	6	7	13	14	15	15	9	14	9	13	12	10	15	10	11	10	12	5
9	16	15	16	14	17	11	18	18	6	11	7	14	7	18	20	16	17	10	19	17	14	17	8	19	12	13	16	12	13
10	15	14	15	7	15	10	15	17	6	6	11	10	8	14	16	13	17	10	16	14	17	11	11	17	12	13	12	15	7
8	14	11	12	10	12	8	14	12	5	11	9	7	11	9	11	12	12	9	13	10	13	13	8	14	13	7	12	13	9
4	9	7	9	7	7	6	10	8	4	8	9	8	4	7	10	8	7	6	9	9	10	9	8	10	10	9	10	10	8

PROCESS OF ITERATION

-0.99649	1.08420	-0.38939	0.86397	-0.48006	0.40674	-0.20195	0.64848	1.59293	-1.84142
-0.93291	-0.57147	-0.67652	-0.89443	0.30816	0.59444	0.30813	0.39568	-0.65290	0.89722
0.19947	0.34940	-0.19072	-1.00588	0.47824	-0.50192	-0.20189	-0.03867	-0.09280	-1.08971
-1.69644	-1.45444	-1.24177	-1.04546	-0.86383	-0.69200	-0.52799	-0.36954	-0.21511	-0.06316
0.08757	0.23829	0.39022	0.54470	0.70312	0.86622	1.03824	1.21980	1.41587	1.63113

-0.98820	1.07477	-0.38699	0.85410	-0.47616	0.40145	-0.20106	0.64558	1.57715	-1.82354
-0.92477	-0.56676	-0.67078	-0.88733	0.30376	0.58925	0.30376	0.39055	-0.64745	0.88991
0.19609	0.34706	-0.19024	-1.00106	0.47817	-0.49866	-0.20106	-0.03956	-0.09323	-1.07937
-1.69588	-1.45449	-1.23984	-1.04417	-0.86239	-0.69093	-0.52714	-0.36899	-0.21478	-0.06310
0.08735	0.23780	0.38948	0.54366	0.70180	0.86556	1.03701	1.21879	1.41448	1.62917

-0.98796	1.07449	-0.38690	0.85387	-0.47605	0.40132	-0.20103	0.64539	1.57685	-1.82320
-0.92455	-0.56663	-0.67061	-0.88711	0.30365	0.58907	0.30365	0.39042	-0.64729	0.88967
0.19601	0.34695	-0.19021	-1.00081	0.47802	-0.49855	-0.20103	-0.03957	-0.09322	-1.07912
-1.69580	-1.45443	-1.23979	-1.04413	-0.86235	-0.69089	-0.52712	-0.36897	-0.21477	-0.06310
0.08735	0.23779	0.38945	0.54363	0.70176	0.86552	1.03696	1.21873	1.41441	1.62910

-0.98795	1.07448	-0.38690	0.85386	-0.47604	0.40132	-0.20103	0.64538	1.57681	-1.82318
-0.92454	-0.56662	-0.67061	-0.88710	0.30365	0.58907	0.30365	0.39042	-0.64729	0.88966
0.19601	0.34694	-0.19021	-1.00080	0.47802	-0.49854	-0.20103	-0.03957	-0.09322	-1.07910
-1.69580	-1.45442	-1.23979	-1.04412	-0.86235	-0.69089	-0.52712	-0.36897	-0.21477	-0.06310
0.08735	0.23779	0.38945	0.54363	0.70176	0.86552	1.03696	1.21873	1.41441	1.62909

CONVERGED.

-0.98795	1.07448	-0.38690	0.85386	-0.47604	0.40132	-0.20103	0.64538	1.57681	-1.82318
-0.92454	-0.56662	-0.67061	-0.88710	0.30365	0.58906	0.30365	0.39042	-0.64729	0.88966
0.19601	0.34694	-0.19021	-1.00080	0.47802	-0.49854	-0.20103	-0.03957	-0.09322	-1.07910
-1.69580	-1.45442	-1.23979	-1.04412	-0.86235	-0.69089	-0.52712	-0.36897	-0.21477	-0.06310
0.08735	0.23779	0.38945	0.54363	0.70176	0.86552	1.03696	1.21873	1.41441	1.62909

-0.98795	1.07448	-0.38690	0.85386	-0.47604	0.40132	-0.20103	0.64538	1.57681	-1.82318
-0.92454	-0.56662	-0.67061	-0.88710	0.30365	0.58906	0.30365	0.39042	-0.64729	0.88966
0.19601	0.34694	-0.19021	-1.00080	0.47802	-0.49854	-0.20103	-0.03957	-0.09322	-1.07910
-1.69580	-1.45442	-1.23979	-1.04412	-0.86235	-0.69089	-0.52712	-0.36897	-0.21477	-0.06310
0.08735	0.23779	0.38945	0.54363	0.70176	0.86552	1.03696	1.21873	1.41441	1.62909

RESULT OF ITERATION

ABILITY OF SCORE GROUP

(1)	-1.69580
(2)	-1.45442
(3)	-1.23979
(4)	-1.04412
(5)	-0.86235
(6)	-0.69089
(7)	-0.52712
(8)	-0.36897
(9)	-0.21477
(10)	-0.06310
(11)	0.08735
(12)	0.23779
(13)	0.38945
(14)	0.54363
(15)	0.70176
(16)	0.86552
(17)	1.03696
(18)	1.21873
(19)	1.41441
(20)	1.62909

EASINESS OF ITEMS

(1)	-0.98795
(2)	1.07448
(3)	-0.38690
(4)	0.85386
(5)	-0.47604
(6)	0.40132
(7)	-0.20103
(8)	0.64538
(9)	1.57681
(10)	-1.82318
(11)	-0.92454
(12)	-0.56662
(13)	-0.67061
(14)	-0.88710
(15)	0.30365
(16)	0.58906
(17)	0.30365
(18)	0.39042
(19)	-0.64729
(20)	0.88966
(21)	0.19601
(22)	0.34694
(23)	-0.19021
(24)	-1.00080
(25)	0.47802
(26)	-0.49854
(27)	-0.20103
(28)	-0.03957
(29)	-0.09322
(30)	-1.07910

CHISQUARE FOR INDIVIDUAL ITEM

(1)	21.41161
(2)	13.81233
(3)	17.07592
(4)	15.80180
(5)	18.25840
(6)	13.72862
(7)	23.17893
(8)	12.79935
(9)	27.04062
(10)	12.96510
(11)	46.92620
(12)	40.54741
(13)	19.06627
(14)	31.83069
(15)	19.57032
(16)	18.99052
(17)	9.19450
(18)	23.92085
(19)	15.71410
(20)	15.05693
(21)	18.33378
(22)	24.62382
(23)	14.93631
(24)	24.93752
(25)	25.26431
(26)	24.27754
(27)	19.84907
(28)	16.52374
(29)	20.46292
(30)	23.81055

CHISQUARE FOR WHOLE TEST= 629.91002

	SLOPE FOR EACH ITEM	TEST OF UNIT SLOPE
(1)	0.99305	-0.04563
(2)	0.86716	-1.27052
(3)	0.85193	-1.52811
(4)	0.88259	-1.13807
(5)	0.93313	-0.57549
(6)	1.08015	0.61518
(7)	1.17201	0.92385
(8)	0.74746	-3.49206
(9)	0.91900	-0.37400
(10)	0.81790	-1.47455
(11)	0.61081	-2.75042
(12)	0.70409	-2.23256
(13)	0.87581	-1.14952
(14)	0.88268	-0.70859
(15)	0.75991	-2.31088
(16)	0.70352	-4.98817
(17)	0.77771	-2.93998
(18)	0.71884	-3.41859
(19)	0.97149	-0.25549
(20)	0.80111	-2.21968
(21)	0.75361	-2.75879
(22)	0.66486	-5.35718
(23)	0.68128	-5.36712
(24)	0.84124	-1.90673
(25)	0.56253	-9.47080
(26)	0.66668	-5.95954
(27)	0.83603	-1.63686
(28)	0.72564	-3.93181
(29)	0.64737	-4.81485
(30)	0.67189	-4.53136

EASINESS	STANDARD ERROR OF MEASUREMENT
-0.98795	0.01859
1.07448	0.02011
-0.38690	0.01472
0.85386	0.01782
-0.47604	0.01506
0.40132	0.01497
-0.20103	0.01427
0.64538	0.01623
1.57681	0.02829
-1.82318	0.03287
-0.92454	0.01799
-0.56662	0.01547
-0.67061	0.01606
-0.88710	0.01766
0.30365	0.01463
0.58906	0.01588
0.30365	0.01463
0.39042	0.01492
-0.64729	0.01592
0.88966	0.01815
0.19601	0.01436
0.34694	0.01477
-0.19021	0.01426
-1.00080	0.01872
0.47802	0.01530
-0.49854	0.01515
-0.20103	0.01427
-0.03957	0.01414
-0.09322	0.01416
-1.07910	0.01956

ABILITY	STANDARD ERROR OF MEASUREMENT
-1.69580	0.03045
-1.45442	0.01873
-1.23979	0.01433
-1.04412	0.01084
-0.86235	0.00985
-0.69089	0.00729
-0.52712	0.00571
-0.36897	0.00442
-0.21477	0.00495
-0.06310	0.00535
0.08735	0.00551
0.23779	0.00575
0.38945	0.00688
0.54363	0.00861
0.70176	0.00816
0.86552	0.01138
1.03696	0.00984
1.21873	0.01273
1.41441	0.01737
1.62909	0.02805

CALIBRATION OF PERSON ABILITY

SUBJECT	TOTAL SCORE	SCORE GROUP	ABILITY
(1)	27.0000	SCORE NOT IN THE CALIBRATED GROUP	
(2)	23.0000	23.0000	1.4144
(3)	22.0000	22.0000	1.2187
(4)	18.0000	18.0000	0.5436
(5)	19.0000	19.0000	0.7018
(6)	23.0000	23.0000	1.4144
(7)	19.0000	19.0000	0.7018
(8)	24.0000	24.0000	1.6291
(9)	20.0000	20.0000	0.8655
(10)	22.0000	22.0000	1.2187
(11)	16.0000	16.0000	0.2378
(12)	19.0000	19.0000	0.7018
(13)	16.0000	16.0000	0.2378
(14)	16.0000	16.0000	0.2378
(15)	18.0000	18.0000	0.5436
(16)	16.0000	16.0000	0.2378
(17)	18.0000	18.0000	0.5436
(18)	14.0000	14.0000	-0.0631
(19)	19.0000	19.0000	0.7018
(20)	14.0000	14.0000	-0.0631
(21)	14.0000	14.0000	-0.0631
(22)	12.0000	12.0000	-0.3690
(23)	16.0000	16.0000	0.2378
(24)	17.0000	17.0000	0.3895
(25)	14.0000	14.0000	-0.0631
(26)	14.0000	14.0000	-0.0631
(27)	15.0000	15.0000	0.0873
(28)	9.0000	9.0000	-0.8623

(4) Prgogramme of conventional comparison of item easiness as calibrated from the sample and different groups

```
      TRACE 0
      READ FROM(CR)
      TRACE 2
      MASTER MAK CHIU
C     PROGRAMME TO FIND
C     MEAN AND STANDARD DEVIATION
C     CORRELATION COEFFICIENT
C     CORRELATED T-TEST OR NONCORRELATED T-TEST
C     HARTLYS TEST
      DIMENSION X(50),Y(50),Z(50)
      READ(1,10) N
10    FORMAT(I2)
      READ(1,20)(X(I),I=1,N)
20    FORMAT(F8.5)
      READ(1,30)(Y(I),I=1,N)
30    FORMAT(F8.5)
      READ(1,31)(Z(I),I=1,N)
31    FORMAT(F8.5)
      CALL XTD(X,N,BARX,STDX)
      CALL XTD(Y,N,BARY,STDY)
      CALL XTD(Z,N,BARZ,STDZ)
      CALL CORR(X,Y,N,RA)
      CALL CORR(X,Z,N,RB)
      CALL CORR(Y,Z,N,RC)
      CALL HART(STDX,STDY,N,HARTLYA,NDFA)
      CALL HART(STDX,STDZ,N,HARTLYB,NDFB)
      CALL HART(STDY,STDZ,N,HARTLYC,NDFC)
      CALL XNOCRT(X,Y,N,TNA,IDFA)
      CALL XNOCRT(X,Z,N,TNB,IDFB)
      CALL XNOCRT(Y,Z,N,TNC,IDFC)
      WRITE(2,40)
40    FORMAT(//////////12X,53HITEM EASINESS        ITEM EASINESS        IT
     1EM EASINESS)
      WRITE(2,41)
41    FORMAT(12X,55HCALIBRATED FROM     CALIBRATED FROM     CALIBRATED F
     2ROM)
```

```
      WRITE(2,42)
42    FORMAT(12X,55HTOTAL SAMPLE          HIGH SCORE GROUP    LOW SCORE GR
     3OUP)
      WRITE(2,33)(X(I),Y(I),Z(I),I=1,N)
33    FORMAT(12X,F10.5,10X,F10.5,10X,F10.5)
      WRITE(2,43) BARX,BARY,BARZ
43    FORMAT(////8X,4HMEAN,2X,F10.5,5X,F10.5,5X,F10.5)
      WRITE(2,44)
44    FORMAT(//1X,9HSTANDARD-)
      WRITE(2,45) STDX,STDY,STDZ
45    FORMAT(2X,9HDEVIATION,2X,F10.5,5X,F10.5,5X,F10.5)
      WRITE(2,51)
51    FORMAT(////12X,61HTOTAL SAMPLE V.S.   TOTAL SAMPLE V.S.   HIGH SCO
     4RE GROUP V.S.)
      WRITE(2,52)
52    FORMAT(12X,55HHIGH SCORE GROUP    LOW SCORE GROUP    LOW SCORE GR
     5OUP)
      WRITE(2,53) RA,RB,RC
53    FORMAT(////1X,11HCORRELATION,/1X,11HCOEFFICIENT,2X,F10.5,6X,F10.5,
     65X,F10.5)
      WRITE(2,54) TNA,TNB,TNC
54    FORMAT(////5X,7HT VALUE,2X,F10.5,5X,F10.5,5X,F10.5)
      WRITE(2,55) IDFA,IDFB,IDFC
55    FORMAT(///2X,11HDEGREE OF  ,/5X,7HFREEDOM,2X,I5,15X,I5,15X,I5)
      WRITE(2,56) HARTLYA,HARTLYB,HARTLYC
56    FORMAT(////1X,14HHARTLY TEST   ,F10.5,5X,F10.5,5X,F10.5)
      WRITE(2,57) NDFA,NDFB,NDFC
57    FORMAT(//2X,9HDEGREE OF,/5X,7HFREEDOM,2X,I5,15X,I5,15X,I5)
      STOP
      END
```

```
      SUBROUTINE CORR(X,Y,N,R)
      DIMENSION X(50),Y(50)
      A=0.0
      B=0.0
      C=0.0
      D=0.0
      E=0.0
      DO 200 I=1,N
      A=A+X(I)*Y(I)
      B=B+X(I)
      C=C+Y(I)
      D=D+(X(I))**2
      E=E+(Y(I))**2
200   CONTINUE
      F=B**2
      G=C**2
      SN=N
      XR=SQRT((D-F/SN)*(E-G/SN))
      R=(A-B*C/SN)/XR
      RETURN
      END
```

```
      SUBROUTINE HART(STDX,STDY,N,HARTLY,NDF)
      VARX=(STDX)**2
      VARY=(STDY)**2
      XN=N
      IF(VARX-VARY) 1,1,2
1     HARTLY=VARY/VARX
      GO TO 3
2     HARTLY=VARX/VARY
3     XDF=2*XN-2.0
      NDF=XDF
      RETURN
      END
```

```
      SUBROUTINE XNOCRT(X,Y,N,TN,IDF)
      DIMENSION X(50),Y(50)
      A=0.0
      B=0.0
      AA=0.0
      BB=0.0
      XN=N
      DO 100 I=1,N
      A=A+X(I)
      AA=AA+X(I)**2
      B=B+Y(I)
      BB=BB+Y(I)**2
100   CONTINUE
      BARX=A/XN
      BARY=B/XN
      A=(A**2)/XN
      B=(B**2)/XN
      SSX=AA-A
      SSY=BB-B
      SSXY=(SSX+SSY)/(2.0*XN-2.0)
      SSPL=SQRT(2.0*SSXY/XN)
      TN=(BARX-BARY)/SSPL
      DF=2.0*XN-2.0
      IDF=DF
      RETURN
      END
```

```
      SUBROUTINE XTD(X,N,BAR,STD)
      DIMENSION X(50)
      C=0.0
      CC=0.0
      DO 100 I=1,N
      C=C+X(I)
      CC=CC+X(I)**2
100   CONTINUE
      XN=N
      BAR=C/XN
      STD=(CC-(C**2)/XN)/(XN-1.0)
      STD=SQRT(STD)
      RETURN
      END
```

```
PROGRAM JUCC
EXTENDED DATA (22AM)
COMPACT PROGRAM (DBM)
CORE                5696

                 SEG      MAKCHIU
                 SEG      XTD
                 SEG      CORR
                 SEG      HART
                 SEG      XNOCRT
                 SEG      SORT
                 ENT      FTRAP
                 ENT      FRESET
```

(5) Programme of conventional comparison of person ability as calibrated from different test subsets

```
      TRACE 0
      READ FROM(CR)
      TRACE 2
      MASTER MAK CHIU
C     PROGRAMME OF STATISTICAL ANALYSIS OF ABILITY CALIBRATION
C     CORRELATED T TEST AND MEAN AND STANDARD DEVIATION
      COMMON X(500),Y(500)
      READ(1,10) N
10    FORMAT(I3)
      READ(1,20)(X(I),I=1,N)
20    FORMAT(F7.4)
      READ(1,30)(Y(I),I=1,N)
30    FORMAT(5F7.4)
      WRITE(2,40)
40    FORMAT(////10X,65HPERSON ABILITY CALIBRATED FROM      PERSON ABILIT
     1Y CALIBRATED FROM)
      WRITE(2,41)
41    FORMAT(10X,44HHARD TEST                                  EASY TEST)
      DO 100 I=1,N
      NO=I
      IF(X(I).EQ.0.0) GO TO 1
      IF(Y(I).EQ.0.0) GO TO 2
      WRITE(2,42) NO,X(I),Y(I)
42    FORMAT(5X,1H(,I3,1H),F10.4,25X,F10.4)
      GO TO 100
1     Y(I)=0.0
      GO TO 3
2     X(I)=0.0
3     WRITE(2,43) NO
43    FORMAT(5X,1H(,I3,1H),38H**********ABILITY CANNOT BE CALIBRATED)
100   CONTINUE
      CALL XTD(X,N,BARX,STDX)
      CALL XTD(Y,N,BARY,STDY)
      CALL CRTT(N,S,T,MDF)
      CALL CORR(N,R)
      WRITE(2,50) BARX,BARY
```

```
50    FORMAT(////1X,6HMEAN =,3X,F10.5,25X,F10.5)
      WRITE(2,53) STDX,STDY
53    FORMAT(//1X,9HSTANDARD-,/1X,9HDEVIATION,F10.5,25X,F10.5)
      WRITE(2,51) T
51    FORMAT(///2X,19HCORRELATED T TEST =,F10.5)
      WRITE(2,54) MDF
54    FORMAT(2X,24HWITH DEGREE OF FREEDOM =,I5)
      WRITE(2,55) R
55    FORMAT(/////2X,27HCORRELATION COEFFICIENT IS=,F10.5)
      STOP
      END




      SUBROUTINE CORR(N,R)
      COMMON X(500),Y(500)
      M=0
      DO 300 I=1,N
      IF(X(I).EQ.0.0) GO TO 300
      IF(Y(I).EQ.0.0) GO TO 300
      M=M+1
300   CONTINUE
      A=0.0
      B=0.0
      C=0.0
      D=0.0
      E=0.0
      DO 200 I=1,N
      A=A+X(I)*Y(I)
      B=B+X(I)
      C=C+Y(I)
      D=D+(X(I))**2
      E=E+(Y(I))**2
200   CONTINUE
      F=B**2
      G=C**2
      SN=M
      XR=SQRT((D-F/SN)*(E-G/SN))
      R=(A-B*C/SN)/XR
      RETURN
      END
```

```
      SUBROUTINE CRTT(N,S,T,MDF)
      DIMENSION D(500)
      COMMON X(500),Y(500)
      S=N
      DO 100 I=1,N
      IF(X(I).EQ.0.0) GO TO 1
      IF(Y(I).EQ.0.0) GO TO 2
      GO TO 100
1     Y(I)=0.0
      S=S-1.0
      GO TO 100
2     X(I)=0.0
      S=S-1.0
100   CONTINUE
      DO 200 I=1,N
      D(I)=X(I)-Y(I)
200   CONTINUE
      A=0.0
      B=0.0
      DO 300 I=1,N
      A=A+D(I)
      B=B+D(I)**2
300   CONTINUE
      TEMP=(S*B-A**2)/(S-1.0)
      TEMP=SQRT(TEMP)
      T=A/TEMP
      DF=S-1.0
      MDF=DF
      RETURN
      END
```

```
      SUBROUTINE XTD(X,N,BAR,STD)
      DIMENSION X(500)
      M=0
      DO 300 I=1,N
      IF(X(I).EQ.0.0) GO TO 300
      M=M+1
300   CONTINUE
      C=0.0
      CC=0.0
      DO 100 I=1,N
      C=C+X(I)
      CC=CC+X(I)**2
100   CONTINUE
      XN=M
      BAR=C/XN
      STD=(CC-(C**2)/XN)/(XN-1.0)
      STD=SQRT(STD)
      RETURN
      END
```

```
PROGRAM JUCC
EXTENDED DATA (22AH)
COMPACT PROGRAM (DBM)
CORE             8192

             SEG      MAKCHIU
             SEG      XTD
             SEG      CRTT
             SEG      CORR
             SEG      SQRT
             ENT      FTRAP
             ENT      FRESET
```

(6) Programme of calculating the standardized difference scores

```
      TRACE 0
      READ FROM(CR)
      TRACE 2
      MASTER MAK CHIU
C     PROGRAMME TO FIND THE STANDARDIZED DIFFERENCE SCORE
C     A FOR HARD
C     B FOR EASY
C    SEA IS MEASUREMENT ERROR ASSOC WITH THE AKEY(PARAMETER CALIBRATED)
C    SEB IS MEASUREMENT ERROR ASSOC WITH THE BKEY(PARAMETER CALIBRATED)
      COMMON X(500),Y(500),SEA(50),SEB(50),AKEY(50),BKEY(50)
      READ(1,10) N,NKA,NKB
10    FORMAT(3I3)
      READ(1,21)(X(I),I=1,N)
      READ(1,20)(Y(I),I=1,N)
      READ(1,30)(AKEY(K),K=1,NKA)
      READ(1,30)(BKEY(K),K=1,NKB)
      READ(1,40)(SEA(K),K=1,NKA)
      READ(1,40)(SEB(K),K=1,NKB)
21    FORMAT(F7.4)
20    FORMAT(5F7.4)
30    FORMAT(F7.4)
40    FORMAT(F7.5)
      CALL STNSR(N,NKA,NKB,BAR,STD,AVX,AVY)
      WRITE(2,60) BAR,STD
60    FORMAT(///2X,39HMEAN OF STANDARDIZED DIFFERENCE SCORE =,F10.5,
     1//2X,7HSTD IS=,F10.5)
      WRITE(2,50) AVX,AVY
50    FORMAT(/////2X,35HMEAN OF MEASUREMENT ERROR OF SET 1=,F10.5,
     2//2X,35HMEAN OF MEASUREMENT ERROR OF SET 2=,F10.5)
      STOP
      END
```

```
      SUBROUTINE STNSR(N,NKA,NKB,BAR,STD,AVX,AVY)
      DIMENSION D(500)
      COMMON X(500),Y(500),SEA(50),SEB(50),AKEY(50),BKEY(50)
      M=0
      AVX=0.0
      AVY=0.0
      DO 100 I=1,N
      IF(X(I).EQ.0.0) GO TO 1
      IF(Y(I).EQ.0.0) GO TO 1
      D(I)=X(I)-Y(I)
      SEX=0.0
      DO 110 K=1,NKA
      IF(X(I).EQ.AKEY(K)) GO TO 2
      GO TO 110
2     SEX=SEA(K)
110   CONTINUE
      IF(SEX.EQ.0.0) GO TO 3
      GO TO 4
3     MO=I
      WRITE(2,10) MO
10    FORMAT(///2X,30HERROR IN CARD READ IN( X DATA),I5)
4     SEY=0.0
      DO 120 K=1,NKB
      IF(Y(I).EQ.BKEY(K)) GO TO 5
      GO TO 120
5     SEY=SEB(K)
120   CONTINUE
      IF(SEY.EQ.0.0) GO TO 6
      AVX=AVX+SEX
      AVY=AVY+SEY
      D(I)=D(I)/(SQRT(SEX+SEY))
      M=M+1
      GO TO 100
6     MO=I
      WRITE(2,20) MO
20    FORMAT(///2X,30HERROR IN CARD READ IN( Y DATA),I5)
```

```
      GO TO 100
1     D(I)=0.0
100   CONTINUE
      A=0.0
      AA=0.0
      DO 200 I=1,N
      A=A+D(I)
      AA=AA+D(I)**2
200   CONTINUE
      XM=M
      AVX=AVX/XM
      AVY=AVY/XM
      BAR=A/XM
      STD=(AA-(A**2)/XM)/(XM-1.0)
      STD=SQRT(STD)
      WRITE(2,30)
30    FORMAT(///2X,30HSTANDARDIZED DIFFERENCE SCORES)
      DO 300 I=1,N
      IF(D(I).EQ.0.0) GO TO 7
      WRITE(2,40) D(I)
40    FORMAT(F10.6)
      GO TO 300
7     WRITE(2,41)
41    FORMAT(2X,40H**********  S.D.SCORE CANNOT BE OBTAINED)
300   CONTINUE
      RETURN
      END


PROGRAM JUCC
EXTENDED DATA (22AM)
COMPACT PROGRAM (DBM)
CORE          8320


                SEG     MAKCHIU
                SEG     STNSR
                SEG     SQRT
                ENT     FTRAP
                ENT     FRESET
```